Not to be cited without
permission of the authors[1]

Canadian Atlantic Fisheries
Scientific Advisory Committee

CAFSAC Research Document 91/15

Ne pas citer sans
autorisation des auteurs[1]

Comité scientifique consultatif des
pêches canadiennes dans l'Atlantique

CSCPCA Document de recherche 91/15

FORECASTING PRESEASON AND INSEASON ATLANTIC SALMON RETURNS TO THE MIRAMICHI RIVER: PARAMETRIC AND NON-PARAMETRIC APPROACHES

by

R.R. Claytor, R.G. Randall, and G.J. Chaput
Department of Fisheries & Oceans
Science Branch, Gulf Region
P.O. Box 5030
Moncton, New Brunswick
E1C 9B6

[1]This series documents the scientific basis for fisheries management advice in Atlantic Canada. As such, it addresses the issues of the day in the time frames required and the Research Documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research Documents are produced in the official language in which they are provided to the Secretariat by the author.

[1]Cette série documente les bases scientifiques des conseils de gestion des pêches sur la côte atlantique du Canada. Comme telle, elle couvre les problèmes actuels selon les échéanciers voulus et les Documents de recherche qu'elle contient ne doivent pas être considérés comme des énoncés finals sur les sujets traités mais plutôt comme des rapports d'étape sur les études en cours.

Les Documents de recherche sont publiés dans la langue officielle utilisée par les auteurs dans le manuscrit envoyé au secrétariat.

Abstract

Three model types, linear regression, time-series, and non-parametric probability distribution models, were compared for their ability to accurately forecast pre-season returns of multi-sea-winter returns of Atlantic salmon (Salmo salar) to the Miramichi River. A jackknife procedure and the number of years the forecast was within ±30% were the criteria used to provide an objective method of judging a model's accuracy. Probability models were judged to be the most effective for pre-season forecasts. A procedure based on probability distribution models was proposed for in-season forecasting.

Résumé

On a comparé trois types de modèle (à régression linéaire, à séries chronologiques et à répartition non paramétrique des probabilités) dans le but de déterminer lequel était susceptible de fournir des prévisions exactes des remontées pré-saisonnières de saumons l'Atlantique (Salmo salar) redibermarins dans la rivière Miramichi, La méthode de rééchantillonnage (jackknife) et le nombre d'années durant lesquelles la prévision était exacte à ± 30 % près ont été les critères utilisés pour la sélection d'une méthode objective d'évaluation de l'exactitude d'un modèle. Les modèles fondés sur les probabilités se sont avérés les plus efficaces pour les prévisions pré-saisonnières. En ce qui concerne les prévisions de remontées durant la saison, on a proposé l'adoption d'une méthode reposant sur des modèles de répartition des probabilités.

Introduction

Forecasting returns of multi-sea-winter (MSW) Atlantic salmon (Salmo salar) one year in advance is an important part of the annual assessment of Atlantic salmon in the Miramichi River. Forecasts of MSW salmon are important because the principal management objective is for egg requirements to be met from these fish and approximately 70% of the egg deposition in the Miramichi River comes from MSW salmon (Randall 1989). If managers know in advance what returns of MSW salmon to expect, and have accompanying estimates of uncertainty, they can set harvest and/or effort levels that have the greatest chance of permitting escapement that will meet spawning requirements. In past assessments (Randall and Schofield 1987; 1988; and Randall et al. 1989a; 1990) linear regression models have been used to forecast MSW returns to the Miramichi River. Unfortunately, models developed in one year were not always applicable in subsequent years and ensuing changes in the models eroded their utility for forecasting returns.

The first part of this document examines two alternatives to linear regression models for preseason forecasting, time series and probability distribution models. Time series modelling has been applied to stock recruitment questions (Noakes et al. 1987), forecasting fishery harvests (Stergiou 1989, Mendelssohn and Cury 1989), species interactions (Stone and Cohen 1990) and environmental effects on Pacific salmon catches (Quinn and Marshall 1989). The two probability distribution models considered have had previous applications in forecasting using the Cauchy (Evans and Rice 1988; Rice and Evans 1988) and the Gaussian kernel estimators (Noakes 1989).

This first section begins by reviewing the results from previous regression models, followed by the results from time series and probability distribution models. This section concludes with a proposed framework of objective criteria for determining which of the possible models is the most appropriate.

The second section discusses a possible model for providing inseason forecasts. This method, outlined by Noakes (1989), uses the Gaussian kernel estimator. This method may allow managers to judge the inherent risk in adhering to preseason management decisions as a season progresses. Inseason forecasts would provide managers with additional opportunities for adjusting preseason regulations that could hedge against unexpected low returns or take advantage of higher than expected returns.

1. PRESEASON MODELS

Data Description.-Preseason models forecast MSW salmon (year i) returns using the numbers of 1SW salmon (year i-1). Returning 1SW salmon provide the first measure, although indirect, of survival from the smolt class which will produce the expected MSW returns. Estimates of these returns to the Miramichi River are obtained from Millbank trap counts made from 1971-1989 (Randall et al. 1990). The data used in the models are trap counts, catch per unit effort, and total returns estimated from trap catch efficiencies (Tables 1,2). Other variables have also been used to explain residuals from the 1SW - MSW models to improve forecasts. These variables are: the proportion of 1SW salmon which were females (Table 1); the commercial catches of small salmon (year i-1) in Salmon Fishing Areas 2 and 4 (SFA2 and SFA4) of the Newfoundland commercial fishery (O'Connell et al. 1990); and the numbers of North American salmon of river age 3 or less caught in the Greenland commerical fishery (year i-1) (GREEN) (Table 1). Tag returns and scale analyses indicate that these fisheries harvest salmon of Miramichi origin (Saunders 1969; Ruggles and Ritter 1980; Pippy 1982). As a result, catches in these fisheries may be expected to influence returns to the Miramichi River. The proportion of female 1SW salmon was used as an indirect measure of the proportion of each smolt class that matured after one year at sea (and assumed age at maturity was environmentally determined) (Marshall et al. 1982).

Regression Models

The original forecast model.-The original forecast model used for predicting MSW salmon returns included two independent variables:

$$(1)\quad MSW_{(yr\ i)}=1SW_{(yr\ i-1)}+PF_{(yr\ i-1)}$$

where MSW was the number of MSW salmon returns in year i, 1SW was the number of 1SW salmon returns in year i-1, and PF was the proportion of 1SW salmon (year i-1) which were females (arcsin transformed).

Total MSW and 1SW returns were estimated by dividing Millbank trap counts by trap efficiencies and adding harvests below Millbank (Table 1). Trap catch efficiencies were estimated using mark-recapture data from 1973 (Turner 1983) and from 1985 to 1987 (Randall et al. 1989a). Efficiencies from 1981 to 1984 were estimated by Randall et al. (1989a). Trap efficiencies were different in these years because of habitat disruption from major dredging activities in the Miramichi estuary beginning in 1981 (Marshall et al. 1982).

Returns in 1987 and 1988 (shown below) were substantially less than returns predicted from equation (1) and refinements to the forecast model were subsequently made.

	Forecast	Returns
1987	54170	19421
1988	36378	21745

Refinements fo the Forecast Model.-The first refinement was to make adjustments for annual variations in fishing effort at the Millbank trap. Because of reductions in personnel, Millbank trap was operated for a shorter period of time and was checked less frequently since 1985 compared to earlier years (Table 2). Prior to 1987, the trap was installed as soon as possible after ice-out in spring (early May) and was operated until early to late November. Beginning in 1987, the period of operation was standardized from May 15 to October 15; counts of salmon before and after these dates usually accounted for less than 1% of the total run (Randall and Schofield 1987). Between 1971 and 1989, the number of operating days ranged between 178 (1975) and 144 (1987).

Under normal working conditions, the trap was visited and hauled once daily or possibly twice daily if two slack tides occurred during regular working hours and if weather permitted. Nevertheless, there was a significant reduction in the number of two-visit days in recent years, from an average of about 60 from 1971 to 1982, to an average of about 20 from 1983 to 1989. This change reduced the number of visits per season for the same periods by about 23%. Number of visits per season has ranged between 268 (1977) and 158 (1989).

The number of salmon captured each year was a direct function of the number of trap visits. Catch per visit was similar for one or two visit days and also similar between the first and second visits on two visit days. For all years combined, there was no significant difference between the number of fish caught during the first and second visits for either 1SW or MSW salmon (t-test, $p>0.61$ and $p>0.85$, respectively). Therefore, to standardize Millbank data, counts were divided by visits per year to calculate an annual catch per unit of effort (CPUE) as an index of abundance. For 1SW salmon, annual CPUE ranged between 4.0 (1983) and 22.2 (1976), and for MSW salmon annual CPUE ranged between 0.8 (1981) and 7.8 (1974).

Regression Model 2.-In addition to using CPUE rather than total returns two additional changes were made to the model. First, the data point for 1974 (year of MSW salmon returns) was removed because it was a significant outlier (studentized residual >3; Wilkinson 1989; Neter et al. 1983). Second, 1SW salmon were divided into male and female components which were entered into the regression model seperately. Only male 1SW salmon were significantly correlated to MSW salmon returns and therefore females were removed from the model. The resulting regression model was:

$$(2) \quad CPUE_{MSW(yr\ i)} = CPUE_{1SW(yr\ i-1)}$$

where $CPUE_{MSW(yr\ i)}$ was the CPUE of MSW salmon at Millbank trap in year i, and $CPUE_{1SW(yr\ i-1)}$ was the CPUE of 1SW male salmon in year i-1. This regression model was significant ($F=16.36$, $p<0.001$); however, the coefficient of determination was only 52% (Figure 1). It is also important to note that one data value (1977 MSW salmon) had a very high leverage value (Fig. 1).

Regression Model 3.-Residuals plotted against time (year) in Model 2 above indicated that MSW salmon returned in lower numbers than expected in recent years, particularly in 1981, 1983, 1987, and 1989 (Fig. 1). As noted previously, dredging activities in the estuary below Millbank may have caused a change in the migration routes of salmon in the vicinity of the trap beginning in 1981. To test this hypothesis, an indicator (qualitative) variable was introduced into the model, whereby a value of 0 was used for years prior to 1981, and value of 1 was applied to years since 1981. The resulting model had a higher coefficient of determination (0.78) than the simple model (Equation 2). However, further analysis indicated that this was not an appropriate forecast model. Incorporating the indicator variable into the regression resulted in a two-slope model; the correlation between 1SW and MSW salmon was positive and significant for the earlier years (1971-1980), but there was no correlation in later years. Thus, although this model fitted the data well, it did not have any predictive power for MSW salmon in recent years (a similar prediction would be obtained by using average MSW salmon returns since 1980).

Regression Model 4.-Salmon landings of small and large salmon from all Newfoundland and Labrador areas (Salmon Fishing Areas (SFAs) 1 to 14) for years 1974 and 1988 (O'Connell et al. 1990) were compared to residuals from the 1SW-MSW salmon regression (Equation 2). Significant negative correlations were observed between Miramichi residuals and small salmon landings in SFA 2 and SFA 4 from the 1SW-MSW model (Equation 2) (Table 3, Fig. 2). Landings in SFA 2 and SFA 4 were significantly and positively correlated, which explained the similar results for these two variables when compared to the Miramichi residuals (Table 3). Landings from SFA 2 were significantly correlated to Miramichi residuals even if the last three years of data were dropped from the regression while correlations with SFA 4 were not significant if 1989 was left out of the data set. Most correlations between the Miramichi residuals and large salmon landings from the different areas were not significant; the exceptions were SFAs 11 and 13 which were significant but the correlations were positive

(Table 3). As a result landings of small salmon from SFA 2 was used as a second independent regressor in the forecast model:

$$(3)\quad CPUE_{MSW(yr\ i)} = CPUE_{1SW(yr\ i-1)} + LAB_{(yr\ i-1)}$$

where $LAB_{(yr\ i-1)}$ was the landings in metric tons of small salmon at SFA 2 in year i-1. Other variables in equation 3 are as in equation 2. Data used in this model are given in Table 4. The multiple regression was significant ($R^2=0.69$; $F=13.58$, $p<0.001$), and coefficients for both independent variables were significant (1SW salmon: positive, $p<0.001$; LAB: negative, $p<0.04$). Residual and leverage plots for this regression are given in Fig. 3.

The suitability of using Equation 3 to forecast MSW salmon returns to the Miramichi River was evaluated by comparing forecasts to returns from 1986 to 1989. The 90% confidence interval for each forecast was large, and returns differed from forecasts by -59% to +48% (Table 5).

Time Series Models

Time series modelling and forecasting can be used to analyze data which meet the following conditions (Hoff 1983): a) measurements are taken at equally spaced intervals; b) there are no missing values in the series being modelled; c) the method of measurement and the event being measured are consistent over time; d) enough data is present; e) short to medium term forecasts are required; and f) the time series is stationary in both the mean and the variance.

The following analysis used Box-Jenkins time series methods to model returns of MSW salmon to the Millbank trapnet based exclusively upon returns of MSW salmon in previous years. Counts of 1SW salmon were also modelled based upon patterns of 1SW returns in previous years.

Model Development.-Returns of 1SW and MSW salmon to the Millbank trapnet between 1971 and 1988 were analyzed. Suggested practical minima for data series are 40 to 50 periods of data or 4 to 5 seasons for seasonal data (Hoff 1983).

Consequently, daily counts were aggregated into the following periods for each year: 1) counts from start of fishing to June 15; 2) counts for June 16 to July 15; 3) counts for July 16 to Aug. 15; 4) counts for Aug. 16 to Sept. 15; and 5) counts for Sept. 16 to end of fishing. This aggregation produced five data points per year, for a total of 90 data values (Figs. 4,5). Seasonal patterns of returns could also be examined and used to advantage in providing inseason forecasts of returns.

The stationarity of the variance of the 1SW and MSW salmon time series was examined using mean/range plot analysis (Hoff 1983).

The Box-Jenkins modelling was performed using SAS time series procedures (ETS, procedure ARIMA) (SAS 1986). Identification procedures and model diagnostics are those suggested by Box and Jenkins (1970) and others (Hoff 1983, Wei 1990).

When the original series was transformed to a log-series prior to modelling, the forecasts were back transformed to the original values using the following procedure (SAS 1986):

(4) back-transformed value=exp(forecast + 2 x standard error)

The data aggregations satisfied the requirements of time series models. The mean/range plot of returns of 1SW and MSW salmon indicated that the variance was not stationary for either 1SW or MSW salmon, the straight line trend indicating that the logarithmic transformation was appropriate for stabilizing the variance (Figs. 4,5).

Only one seasonal differencing (period 5) was required to transform the natural logs into a stationary series. By computing the difference between every fifth value successive, the overall trending behaviour was removed for 1SW and MSW salmon was removed.

Model identification procedures and diagnostics suggested the following models for MSW and 1SW salmon returns at Millbank trapnet using the 1971 to 1988 data series.

$$(5)\quad lnMSW_t = lnMSW_{t-5} + \Theta_0 + (1-\Theta_1\beta-\Theta_2\beta^2)(1-\Theta_*\beta^5)\,e_t$$

where $lnMSW_i$=ln of MSW counts at time i
θ_0=trend parameter (mean of the differenced series)
θ_i=regular moving average parameter
θ_*=seasonal moving average parameter
β=backshift operater
e_t=residual, $\epsilon \sim N(0,s^2)$

$$(6) \quad lnGrilse_t = lnGrilse_{t-5} + \phi - (\phi_1\beta - \phi_2\beta^2 - \phi_3\beta^3)(1 - \phi_*\beta^5)\, lnGrilse_t) + e_t$$

ϕ_i=regular autoregressive parameter
ϕ_*=seasonal autoregressive parameter

The model diagnostics for the MSW salmon counts using 1971 to 1988 data were suitable (Table 6). All parameters estimated had coefficients which were significantly different from 0, including the trend parameter. The parameters were uncorrelated thus the model was not overspecified. The residuals were not autocorrelated (Table 6, Fig. 6), thus the model was able to account for the serial correlation. Residuals were also normally distributed (Fig. 7), although the residuals tended to have a larger scatter at smaller predicted values. The index of determination (R^2) for the model was very low, suggesting that the serial variation of the MSW salmon counts between 1971 and 1988 was small compared to the random variation. The significant trend parameter suggested that the counts of MSW salmon were decreasing over time and subsequent overall forecasts (by year) would always be less than the previous forecast.

For 1SW salmon, the 1971 to 1988 data series model diagnostics were similar to those from the MSW salmon model. All estimated parameter coefficients were significantly different from 0 and parameters were uncorrelated. Not all the autocorrelation of the residuals was removed, although attempts to account for the larger autocorrelation value at lag 16 were not successful (Fig. 8). The residuals were normally distributed with a similarly larger scatter at smaller predicted values (Fig. 9). The index of determination for the model was substantially higher than that for MSW salmon although the random error component was large and forecasts unreliable (Table 7).

Suitability of time series models for forecasts.- The models for MSW and 1SW salmon were run sequentially to provide forecasts of returns for 1986 to 1989 using only the previous years in the estimation of the parameter coefficients. In general, the forecasts for MSW salmon were larger than returns although not in all periods. The forecasts for MSW salmon decreased between 1986 and 1989, a result of a significant decreasing trend parameter in the data series. Absolute percent errors for all four estimates were 23%, indicating that the fitted series differed from the observed series by 23%, although the backtransformed values differed by 117% to 582% of forecasts. Forecasts summed for all periods were 140% to 208% of returns. The confidence intervals were excessively wide and provide no reassurance whatsoever in the forecasting ability of the models. The forecast of MSW salmon counts at Millbank using the overall mean (average count in each time period) gave forecasts which had smaller percent errors than the model forecasts and were 111% to 167% of returns (Table 8).

The residual analysis by period indicates those periods for which the unaccounted variation was large. Periods 3 and 4 had high residuals and also displayed decreasing trends since 1971 (Fig. 7). A sinusoidal trend in period 5 indicated that returns to Millbank between mid-September to the end of the trap season were lower than expected between 1975 and 1982, whereas the 1971 to 1973 and 1983 to 1987 returns were equal to forecast or higher than expected (Fig. 7). These trends in the residuals illustrate the extent of the inefficiency of the models to forecast returns.

In spite of the apparently high proportion of explained serial autocorrelation of the 1SW models, the error associated with the forecast relative to the actual values were 4 times that of the MSW salmon models, ranging between 76% and 85%. Backtransformed percent errors were 81% to 579% from 1986 to 1988. Forecasts of 1SW salmon were larger than returns in all years, ranging between 133% to 240% of returns. Forecasts using the overall mean count of 1SW since 1971 underforecasted the total counts, representing between 45% and 67% of actual (Table 9). Trends in the residuals were prevalent in periods 1 and 4 (Fig. 9). Returns of 1SW salmon were not efficiently simulated with the above Box-Jenkins models.

Probability Distribution Models

Two probability distribution models were compared to determine their ability to provide accurate forecasts and useable measures of uncertainty for managers. These models were those based on the Cauchy distribution in the manner described by (Evans and Rice 1988; Rice and Evans 1988) and the Gaussian distribution in the manner described by Noakes (1989).

A major difference between parametric regression and non-parametric probability distribution models is that parametric models require certain assumptions, for example normally distributed errors, to be met for the derived estimates to be robust. In contrast, non-parametric models allow the data to determine the distribution to be analyzed. The probability distribution models we employed require an initial assumption regarding the distribution of the kernel density estimator, but this does not affect the probability distribution determined from the data.

Model Development.- The Cauchy kernel estimator described by Evans and Rice (1988) is provided below:

$$(7)\quad f(x)=1/[1+\left(\frac{x}{h}\right)^2]$$

where x is the difference between the stock size being estimated and each previously observed stock size and h is the smoothing parameter.

The multivariate Gaussian kernel estimator used in this paper is that given by Noakes (1989) and is given below:

$$(8)\quad f(x)=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_1\ldots h_d(2\pi)^{\frac{d}{2}}}\prod_{j=1}^{d}\exp[-\frac{1}{2}\left\{\frac{(x_j-x_{ij})}{h_{ij}}\right\}^2]$$

where d=the number of variables
x=vector of length d which is composed of the reference values for the variables from which the forecasts are made, in our case these are 1SW (year i-1) and MSW (year i) returns
matrix x_{ij}=the previously observed values for the variables entering the model
h=the vector of smoothing parameters, if the smoothing parameters in each model are identical or if one smoothing parameter is used, this vector is a single value.

The differences in the algorithms used to determine these distributions are that the Cauchy algorithm cannot predict recruitments other than those which have been previously observed and so the cummulative distribution or ogive is described by a step function. A second difference is that recruitment or MSW forecast from the Cauchy distribution is determined from the recruitment value that corresponds to 50% of the cummulative probability distribtuion from the step function (Evans and Rice 1988). The forecast value from the Gaussian distribution corresponds to the point of maximum probability in the distribution with stock size or both stock size and commercial catch held constant.

A number of points must be considered in developing these probability distribution models. First, the number of dimensions or data sets that can be included in the model must be considered. This number depends on two factors, the sample size and the importance of tails in the distributions. The sample size required to ensure that the relative mean square error is <0.1 increases dramatically with the number of dimensions. Silverman (1986; pages 93-94) indicates that with one dimension a sample of 4 is required, two dimensions requires 19, and three requires 67. Additionally, as dimensions are added, the importance of the tails of the distributions also increases (Silverman 1986, page 92). In this evaluation we have restricted our models to two dimensions. With this restriction we meet the minimum required sample number of 19 using years 1971-1989.

A second consideration is pre-scaling the data prior to multivariate analyses to avoid extreme differences in spread among the data sets examined. Silverman (1986, page 77) recommends normalizing the data to zero mean and unit covariance to achieve this effect. When this procedure is followed, it is generally not necessary to consider models with more than one smoothing parameter.

A third consideration occurs if the distribution includes non-zero values when these are not possible because of the nature of the data (Silverman 1986, page 29). In these cases it would be better if no weight were given to negative values. One possible solution to this problem is to truncate the distribution; the problem with this solution is that the probability function will no longer integrate to unity and points near zero will not receive sufficient weight (Silverman 1986, page 29). Another solution is to take the logarithm of the data (Noakes 1989). If only the forecast is required, backtransforming by the antilog will suffice. If the distribution is also required the backtransformation should be done as suggested by Silverman (1986, page 30) as follows:

$$f(x)=\left(\frac{1}{x}\right)(g(\log(x)) \quad for\ x>0$$

A fourth problem to consider is the selection of the smoothing parameter. The smoothing parameter is important because it determines the weight given to surrounding data for any given point, x_i. If h is very large, infinity, equal weight will be given to all values of x and the distribution will be over-smoothed (too flat) to provide an accurate description of the distribution. If h is very small, only x values very close to x^i will receive weight and the resulting contribution of the kernel estimators will be very narrow curves and multiple peak distribtuions will occur, the distribution will be under-smoothed.

Two methods, previously used with fisheries data for estimating smoothing paramters were employed, maximum likelihood (Noakes 1989) and least squares validation (Evans and Rice 1988; Rice and Evans 1988). These two methods were used because the likelihood method may be more sensitive to outliers (Silverman 1986, page 54) but has the advantage of quicker calculation than the least squares method.

A fifth matter to consider is the choice of kernel estimator; two estimators, the Cauchy and Gaussian, are compared in this paper.

Model Development.-Given the above criteria to consider in developing a probability model we used the following procedures to make the necessary comparisons in choosing the most appropriate forecast model. With the Gaussian estimator all data were normalized whether raw untransformed or logged data were used. With the Cauchy estimator the raw untransformed data, was used, as described by Evans and Rice (1988). We compared single smoothing parameters to individual smoothing parameters (a separate smoothing parameter for each variable in the model). Individual smoothing parameters were chosen with both data sets in the model. They were chosen by fixing one parameter and varying the second until the maximum probability was found using the likelihood method, or the minimum probability found using the least squares method. Then the second parameter was fixed and the first allowed to vary until the appropriate criteria were met. This procedure was repeated until the best set of smoothing parameters was found. These models were compared to the Cauchy model algorithm described by Evans and Rice (1988) and Rice and Evans (1988). This algorithm does not use normalized or logged data. Each of the models used 1SW salmon, year i-1, to forecast MSW salmon, year i.

The most appropriate model was selected by examining the residual sum of squares (residual=forecast-returns) and the number of years where the forecast was >±30% of returns. Residuals were calculated after forecasts based on log transformations were backcalculated to arithmetic values. Thus, all residuals were calculated on the same scale. All models were tested using a jackknife approach. That is, a forecast was determined for each year by leaving that year out of the data set, putting it back in taking out the next year.

Two additional data sets were examined to determine if they could explain or improve the initial 1SW - MSW forecast. These data sets were; catch of small salmon in the SFA 4 Newfoundland commercial fishery, year i-1, (SFA4) (O'Connell et al. 1990) and the number of 1SW salmon of river age 3 or less caught in the Greenland fishery, year i-1, (GREEN) (D. Reddin, DFO, St. John's, Nfld) (Table 1). SFA4 was chosen because in a preliminary analysis it was the only area that produced a smoothing parameter that reduced the variance in the Cauchy model. GREEN was chosen because of the large number of North American salmon that are exploited in Greenland.

After the best 1SW-MSW forecast model was determined, based on least residual sum of squares. New forecasts using the SFA4 and GREEN data sets were obtained using this model. These forecasts were obtained by subtracting SFA4 or GREEN forecasted residuals from the initial forecast. These new forecasts were examined using the criteria described above to determine the best model. Data for these sets were available only from 1974 to 1989 and comparisons to mean and Cauchy models were restricted to those years.

Suitability for forecasting.-Probability models provided useful forecasts. All models produced better estimates than the mean. Residual sum of squares and number of years when forecasts were >±30% from returns were less with all probability models when compared to the mean (Table 1). Raw data models using least square smoothing parameters were better than those using the likelihood method. The log normalized model using individual smoothing parameters selected by the likelihood method was the worst of the log models; there was no difference in the other log models (Table 10). The Cauchy model was the best in terms of least residual sum of squares, while the log normalized model with a single likelihood smoothing parameter was the best in terms of fewest years >±30% of forecast (Table 10).

Each of these models presents a slightly different perspective of the data. The least squares approach produced the most stable smoothing parameter estimates whether dealing with raw or logged data. There was the least difference between individual and combined smoothing parameter estimates with this approach compared to the likelihood approach (Table 11). The raw normalized distribution appears multimodal (Fig. 10) while the log normalized distribution is multimodal and has a long tail (Fig. 11). The Cauchy distribution is typically presented as a cumulative distribution (Evans and Rice 1988) and thus it is difficult to comment on its modality; however, it is generally steep around the median, suggesting a sharply peaked unimodal distribution (Fig. 12). The raw normalized least square, the log normalized likelihood, and the Cauchy models were given a closer examination to determine the most appropriate forecast model.

Probability models provide superior forecasts to the mean because of their tendency to detect changes in MSW returns at extreme 1SW levels (Fig. 13). Examination of combined residual sum of squares at various stock sizes demonstrates the superiority of probability models at extreme stock sizes (Fig. 14). There is little difference in MSW forecasts between mean and probability models at intermediate stock levels (Fig. 14).

Annual trends in residuals were similar for all models. There were no apparent positive or negative trends over time and variability was less since 1981 relative to previous years (Fig. 15).

The Cauchy model provides the most appropriate forecasting model. It had the lowest residual sum of squares (Table 10) and was the best at forecasting the extremes (Figs. 13,14). Therefore, residuals from the Cauchy model (Table 12) were examined against SFA4 and GREEN data sets to determine if using these data sets would improve the forecast.

The number of 1SW salmon of river age 3 or less caught at Greenland were better at explaining the residuals in the model than the small salmon catch in SFA4. There was almost no improvement in 1SW-MSW Cauchy residuals with SFA4 compared to a 22% improvement using the GREEN data set. The GREEN data set in the model improved the forecast 44% over that obtained for the mean in the 1974-1989 period (Table 10). The improvement using the GREEN data set comes from both the extreme 1SW values and intermediate stock sizes (Fig. 16).

Forecast for 1991.-A 1991 forecast of 26,000 MSW salmon was derived using the Cauchy model this forecast is equal to mean returns from 1971-1990. This model also suggests there is a 30% probability of returns less than spawning requirements (23,000 MSW) in 1991 and a 50% probability that returns will be between 23,000 and 34,000 MSW salmon (Fig. 17). Returns of 34,000 MSW salmon would be 30% greater than the 26,000 forecast.

2. INSEASON FORECASTS

The procedure described by Noakes (1989) for inseason forecasting were those followed, with a few exceptions, in this analysis. Cumulative returns to date (weekly) and numbers of salmon yet to return were used to derive probability distributions using Gaussian kernel estimators to forecast returns expected to arrive. In this analysis, the first exception to Noakes's (1989) methodology was that instead of using the mean and 95% confidence interval of historical returns as the preseason forecast, the forecast derived using the Gaussian 1SW - MSW model and its associated 95% confidence interval, was used as the preseason forecast. An additional exception was that it was not necessary to transform the data to natural logarithms in order to ensure that total forecasted run size and the lower confidence limit would be at least equal to returns to date. This transformation was not necessary because no confidence intervals of the preseason forecasts included values less than zero. Although this transformation would be necessary before proceeding with the analysis of any additional weeks as the lower limit of the confidence intervals approached, or reached zero for most of the years examined by week 25 (Table 13).

The smoothing parameters for these models were calculated as described previously. Preseason and inseason forecasts were combined according to the weighting scheme proposed by Noakes (1989) to produce a total combined forecast. The weighting scheme is described below:

(9) ***Total forecast***$=(w \times$ ***inseason forecast***$)+(1-w) \times$ ***preseason fore***

where w, the weight, is calculated based on the 95% confidence intervals of the inseason and preseason forecasts as below:

(7) W=A/A+B

where A is the length of the 95% confidence interval for the preseason forecast and B is the length of the 95% confidence interval of the inseason yet to return forecast. Thus, as the season progresses B will become smaller than A and the inseason returns will have more weight than the preseason forecast in determining total forecast.

Model Development.-Our purpose is to demonstrate a potential application of Noakes (1989) methodology for managing Miramichi River salmon stocks. As a result, we provide a detailed analysis of the first five weeks of each season for four years. The weeks analyzed begin with standardized week 22 which begins May 28 and end with week 25 which finishes June 24. Recreational seasons, for bright salmon, have historically begun around June 1 and ended September 30. Commercial seasons usually began between June 1 or June 15 and continued to the end of July. Thus, any major management adjustments suggested by inseason returns to the end of June could still have an effect on returns for a large part of the season.

The four years chosen were 1977 and 1979, the two worst years for preseason forecasts, and 1983 and 1986, two of the better years for preseason forecasts (Table 12). Although 1983 was not the best year for preseason forecasts, it was chosen because of the high proportion of early returns in that year, with the hope that it might prove to be a good example of a year when the confidence limit of the inseason forecast would relatively quickly surpass that of the preseason forecast.

Suitability for forecasting.-For each year, there was a reduction in the 95% confidence interval from week 22 to week 25 (Table 13). For 1983, the confidence interval of the preseason forecast was less than the preseason forecast in the first week analyzed (Table 13). Thus, the inseason forecast reduced the preseason forecast residual by half in the first week of the season. For 1986, the inseason forecast 95% confidence interval was reduced below the preseason interval by week 25 (Table 13).

The forecasts for 1977 and 1979 did not improve during the weeks examined in this model. Although the residuals improved slightly for 1979, it was not an appreciable change that would induce management changes (Table 13).

We conclude that inseason forecasting is a promising technique for improving the use of forecasts in salmon management. Probability distributions provide an efficient means of providing these forecasts. Additional analyses need to be conducted to determine if inseason forecasts for years such as 1977 and 1979 can be improved.

Discussion

Probability distribution models provided forecasts that were closer to returns than either regression or time series models. The original regression model for preseason forecasting of Atlantic salmon in the Miramichi River (Equation 1 and Fig.1) explained only 52% of the interannual variation in MSW salmon returns. The final regression model improved forecasts but these were still not satisfactory for management. For example, the 90% prediction interval was ±87% of the predicted value (Table 5).

Several factors may account for the poor forecasts using the Box-Jenkins time series models: 1) the general decreasing trend in numbers of large salmon at Millbank from 1971 to 1988, 2) the small proportion of the serially correlated variance which could be explained by the models, and 3) the commercial fisheries in Miramichi Bay from 1981 to 1983 which undoubtedly impacted on the counts at Millbank, as shown by the low residuals for large salmon for those years in period 3, July 16 to August 15.

Time series modelling may, nevertheless, be useful for removing serial correlation prior to further analysis. An example that became apparent through residual analysis was the effect of the commercial fishery in Miramichi Bay from 1981 to 1983 on returns at Millbank. Another example was the lower than expected returns of large salmon in all periods and the lower returns of grilse in 4 of 5 periods in 1981, which may be partly attributable to the extensive dredging in the Miramichi estuary that year.

Probability models will be most useful for forecasting at extreme 1SW levels or if additional data sets can be included in the model. This second condition may be difficult to satisfy. The Greenland catches, which were most useful, may not always be available in time for preseason forecasts. However, other data sets may prove equally as useful and may be available in time. These include total Greenland catch, Newfoundland commercial catch in other SFAs, and size distribution of 1SW salmon. These will be examined in future analyses.

While probability models are useful, it is important to have a framework for evaluating their appropriateness. A potential framework for doing this has been suggested in this paper. The jackknife approach is important for reducing bias in the forecast estimates and residual sum of squares is a useful measure of the overall performance of the model. The number of years >±30% may be a useful measure for managers. If one model has lower residuals and another fewer years >±30%, a manager could choose between a model which provides forecasts which overall performs better or one in which returns are more likely to be within some critical range. Finally, it will be useful to view predicted versus residual plots and annual residual plots to determine precisely how the model is performing. This procedure would be similar to the diagnostics commonly employed in regression analyses.

One method for employing inseason forecasting in management may be to proceed with management plans based on preseason forecasts until the confidence interval of the inseason forecast becomes less than the preseason forecast. For the examples presented here this would have resulted in improved forecasts by week 22, the beginning of the season, for 1983 and by week 25 for 1986, about midway through the summer recreational and commercial seasons. Additional analyses will determine the general applicability of this model for use in developing Atlantic salmon management plans.

Recommendations.-Additional analyses which may improve the forecasts of probability distributions and that will be investigated in future analyses include the following:

1. Discount outliers greater than one standard deviation from the mean by the inverse of the distance from the mean. This procedure dampens the effect of outliers on the smoothing parameter estimate.

2. Use standardized residuals (residuals/standard deviation) to determine the residual sum of squares. This procedure will dampen the effect of years with very high residuals.

3. Leave the year with high residuals out of the model to determine its effect on model selection.

4. When using log transformed values, plot the forecast versus predicted values to determine if there is any bias in the log estimates and force the line through zero.

5. Put GREEN and SFA4 data in the model first to determine if they are better predictors of MSW salmon returns than 1SW salmon.

6. Examine the effect the increase in proportion of repeat spawners may have on the forecast.

Acknowledgements

The authors thank Jake Rice for introducing them to the topic of probability distributions and Don Noakes, Gloria Nielsen, Russell Millar, Peter Shelton and other members of the Anadromous and Catadromous Subcommittee for comments on various versions of this analysis. Gloria Nielsen and Claude LeBlanc provided valuable comments on previous drafts of the manuscript.

References

Box, G.E.P., and G.M. Jenkins. 1970. Time series analysis: forecasting and control. Holden-Day, San Francisco. 553 p.

Evans, G.T., and J.C. Rice. 1988. Predicting recruitment from stock size without the mediation of a functional relation. J. Cons. int. Explor. Mer. 44:111-122.

Hoff, J.C. 1983. A practical guide to Box-Jenkins forecasting. Lifetime Learning Publications, Belmont, Calif. 316 p.

Marshall, T.L., J.L. Peppar, and E.J. Schofield. 1982 Prediction of 2SW and older Atlantic salmon returning to Millbank trap, Miramichi River, New Brunswick. CAFSAC Res. Doc. 82/51.

Mendelssohn, R. and P. Cury. 1989. Temporal and spatial dynamics of a coastal pelagic species, Sardinella maderensis off the Ivory Coast. Can. J. Fish. Aquat. Sci. 46:1686-1697.

Neter, J., W. Wasserman, and M.H. Kutner. 1983. Applied linear regression models. Irwin, Homewood, Ill. USA.

Noakes, D. D.W. Welch, and M. Stocker. 1987. A time series approach to stock-recruitment analysis: transfer function noise modelling. Natural Resource Modelling 2:213-233.

Noakes, D.J. 1989. A nonparametric approach to generating inseason forecasts of salmon returns. Can. J. Fish. Aquat. Sci. 46:2046-2055.

O'Connell, M.F., J.B. Dempson, E.G.M. Ash, and N.M. Cochrane. 1990. Status of Atlantic salmon (Salmo salar L.) stocks of the Newfoundland Region, 1989. CAFSAC Res. Doc. 90/43.

Pippy, J. 1982. Report of the working group on the interception of mainland salmon in Newfoundland. Can. Man. Rep. Fish. Aquat. Sci. 1654: 187 p.

Quinn II, J.T., and R.P. Marshall. 1989. Time series analysis: Quantifying variability and correlation in SE Alaska salmon catches and environmental data, p. 67-80. In R.J. Beamish and G.A. McFarlane [ed.] Effects of ocean variability on recruitment and an evaluation of parameters used in stock assessment models. Can. Spec. Publ. Fish. Aquat. Sci. 108.

Randall, R.G., and E.J. Schofield. 1987. Status of Atlantic salmon in the Miramichi River, 1986. CAFSAC Res. Doc. 87/5.

Randall, R.G., and E.J. Schofield. 1988. Status of Atlantic salmon in the Miramichi River, 1987. CAFSAC Res. Doc. 88/49.

Randall, R.G. 1989. Effect of sea-age on the reporductive potential of Atlantic salmon (Salmon salar) in Eastern Canada. Can. J. Fish. Aquat. Sci. 46: 2210-2218.

Randall, R.G., P.R. Pickard, and D. Moore. 1989a. Biological assessment of Atlantic salmon in the Miramichi River, 1988. CAFSAC Res. Doc. 89/73.

Randall, R.G., D. Moore, and P.R. Pickard. 1990. Status of Atlantic salmon in the Miramichi River during 1989. CAFSAC Res. Doc. 90/4.

Rice, J.C., G.T. Evans. 1988. Tools for embracing uncertainty in the management of the cod fishery of NAFO divisions 2J+3KL. J. Cons. int. Explor. Mer. 45:73-81.

Ruggles, C.P., and J.A. Ritter. 1980. Review of North American smolt tagging to assess the Atlantic salmon fishery off West Greenland. Rapp. p.-v. Reun. Cons. int. Explor. Mer. 176;82-92.

SAS. 1986. SAS/ETS User's Guide. Version 6, First Ed. SAS Institute Inc., Cary, N.C. 560 p.

Saunders, R.L. 1969. Contributions of salmon from the Northwest Miramichi River, New Brunswick, to various fisheries. J. Fish. Res. Board Can. 26:269-278.

Stergiou, K.I. 1989. Modelling and forecasting the fishery for pilchard (Sardina pilchardus) in Greek waters using ARIMA time-series models. J. Cons. int. Explor. Mer. 46: 16-23.

Stone, J.N., and Y. Cohen. 1990. Changes in species interactions of the Lake Superior fisheries system after the control of sea lamprey as indicated by time series models. Can. J. Fish. Aquat. Sci. 47: 251-261.

Turner, G.E. 1983. An efficiency estimate of the Millbank estuarial sampling trap, Miramichi River, New Brunswick, 1973. CAFSAC Res. Doc. 83/78.

Wei, W.W.S. 1990. Time series analysis: Univariate and multivariate methods. Addison-Wesley Co., Inc., Redwood City, Calif. 478 p.

Wilkinson, L. 1989. Systat: The system for statistics. Evanston, Ill: Systat, Inc.

Silverman, B.W. 1986. Density estimation for statistics and data analysis. (Monographs on Statistics and Applied Probability). Chapman and Hall, London.

Table 1. Estimated total returns of Atlantic salmon to Miramichi River. Returns are calculated as: Millbank trap count/trap efficiency + harvest in the estuary below Millbank. The proportion of 1SW salmon which were females each year is also indicated (PF). These data were used in Equation 1 (see text) for predicting MSW salmon returns one year in advance.

	1SW Salmon				MSW Salmon				Residual Variables	
Year	Estuary	Trap	Efficieny	Returns	Estuary	Trap	Efficiency	Returns	PF	GREEN
1971	0	1962	0.055	35,673	15128	399	0.043	24,407	0.110	n/a
1972	39	2562	0.055	46,275	2282	1151	0.043	29,049	0.220	n/a
1973	0	2450	0.055	44,545	866	1132	0.043	27,192	0.169	n/a
1974	0	4038	0.055	73,418	941	1791	0.043	42,592	0.302	162130
1975	393	3548	0.055	64,902	724	1208	0.043	28,817	0.274	182080
1976	1780	4939	0.055	91,580	871	943	0.043	22,801	0.241	115210
1977	379	1505	0.055	27,743	6865	1934	0.043	51,842	0.228	143040
1978	1232	1268	0.055	24,287	8377	693	0.043	24,493	0.374	92230
1979	5510	2500	0.055	50,965	1659	318	0.043	9,054	0.274	169450
1980	2697	2139	0.055	41,588	10899	1093	0.043	36,318	0.193	141190
1981	1332	2174	0.034	65,273	7137	199	0.022	16,182	0.251	165330
1982	1997	2665	0.034	80,379	12213	408	0.022	30,758	0.295	150710
1983	1360	810	0.034	25,184	16788	245	0.022	27,924	0.292	27490
1984	1	1010	0.034	29,707	1	333	0.022	15,137	0.217	33230
1985	0	912	0.015	60,800	5	311	0.015	20,738	0.228	113890
1986	16	1763	0.015	117,549	18	469	0.015	31,285	0.220	129320
1987	16	1272	0.015	84,816	21	291	0.015	19,421	0.354	133910
1988	52	1828	0.015	121,919	78	325	0.015	21,745	0.218	78580
1989	31	1128	0.015	75,231	78	257	0.015	17,211	0.220	46730

Table 2. Fishing effort for Atlantic salmon at Millbank trap. Effort is given as days of trap operation and numbers of visits (trap hauls) per year.

	Effort					Catch per unit effort		
		Visits per day				1SW		MSW
Year	Days	1	2	3	Total	Both sexes	Males	
1971	155	107	46	2	205	9.57	8.52	1.95
1972	151	76	74	1	227	11.20	8.74	5.07
1973	159	84	73	2	236	10.38	8.63	4.80
1974	173	117	55	1	230	17.56	12.25	7.79
1975	178	106	72	0	250	14.19	10.30	4.83
1976	174	126	48	0	222	22.25	16.87	4.25
1977	164	62	100	2	268	5.62	4.34	7.22
1978	167	110	54	3	229	5.54	3.47	3.03
1979	170	128	40	2	214	11.68	8.48	1.49
1980	177	128	49	0	226	9.47	7.64	4.84
1981	174	111	63	0	237	9.17	6.87	0.84
1982	164	110	53	1	219	12.17	8.58	1.86
1983	168	135	33	0	201	4.03	2.85	1.22
1984	152	120	32	0	184	5.49	4.30	1.81
1985	164	161	3	0	167	5.46	4.22	1.86
1986	158	140	18	0	176	10.02	7.81	2.66
1987	144	124	20	0	164	7.76	5.01	1.77
1988	148	126	21	1	171	10.69	8.36	1.90
1989	147	136	11	0	158	7.14	5.57	1.63

Table 3. Correlations coefficients for comparisons between residuals in the Miramichi forecast model (Equation 2 in text) and small and large salmon landings in Labrador, Newfoundland, and Greenland. NS is not significant.

	Small salmon (year i-1)	Large salmon (year i)
Labrador SFA1	-0.21 NS	+0.14 NS
Labrador SFA2	-0.55 P<0.05	+0.05 NS
Newfoundland SFA3	-0.02 NS	-0.05 NS
Newfoundland SFA4	-0.57 P<0.05	-0.05 NS
Newfoundland SFA5	-0.38 NS	+0.16 NS
Newfoundland SFA6	+0.06 NS	+0.18 NS
Newfoundland SFA7	-0.02 NS	+0.16 NS
Newfoundland SFA8	-0.07 NS	+0.29 NS
Newfoundland SFA9	-0.10 NS	+0.10 NS
Newfoundland SFA10	-0.30 NS	+0.54 P<0.05
Newfoundland SFA11	-0.15 NS	+0.07 NS
Newfoundland SFA13	-0.45 NS	+0.54 P<0.05
Newfoundland SFA14	-0.09 NS	+0.17 NS
Labrador (Gulf)	+0.06 NS	+0.18 NS
Greenland	-0.15 NS	-0.15 NS

Table 4. Data used in Equation 3 (see text) to predict MSW salmon returns. CPUE is catch per unit effort of effort of salmon at Millbank trap (MSW or 1SW) and LAB is landings (t) in SFA 2 of Labrador.

Year (i)	CPUE (MSW,year i)	CPUE (1SW,year i-1)	LAB (year i-1)
1975	4.83	12.25	82
1976	4.25	10.30	134
1977	7.22	16.87	107
1978	3.03	4.34	92
1979	1.49	3.47	28
1980	4.84	8.48	65
1981	0.84	7.64	168
1982	1.86	6.87	204
1983	1.22	8.58	126
1984	1.81	2.85	71
1985	1.86	4.30	32
1986	2.66	4.22	54
1987	1.77	7.81	102
1988	1.90	5.01	143
1989	1.63	8.36	123
1990	.	5.57	79

Table 5. Comparison between predicted and actual MSW salmon returns to Miramichi River, 1986 to 1989. Predicted returns were calculated using Equation 3 (see text). CPUE is the catch per trap visit (haul) of MSW salmon at Millbank. Trap count was calculated assuming an average number of visits per year of 170. Predicted returns were calculated as the trap count divided by a trap efficiency of 0.015. Actual returns are from Randall et al. 1990.

		Predicted (90% prediction interval)			
Year	R^2 (df)	CPUE	Trap count	Forecast	Returns
1986	0.76 (8)	2.1 (0.0,4.4)	357 (0,748)	23800 (0,49867)	31267 (-31)
1987	0.75 (9)	3.1 (1.0,5.1)	527 (170,867)	35133 (11333,57800)	19400 (+45)
1988	0.73 (10)	1.2 (0.0,3.3)	204 (0,561)	13600 (0,37400)	21667 (-59)
1989	0.72 (11)	2.9 (1.0,4.9)	493 (170,833)	32867 (11333,55533)	17133 (+48)
1990	0.69 (12)	2.3 (0.3,4.3)	391 (51,731)	26067 (3400,48733)	n/a

Table 6. Sample diagnostic statistics for the MSW salmon model, 1971 to 1988.

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MU	-0.04093	0.07528	-0.54	0
MA1,1	-0.29930	0.10885	-2.75	1
MA1,2	-0.24629	0.11472	-2.15	2
MA2,1	0.62549	0.10969	5.70	5

Constant Estimate = -0.0409289

Variance Estimate = 1.187975
Std Error Estimate = 1.08994266
AIC = 259.343928
SBC = 269.067195
Number of Residuals= 84

Correlations of the Estimates

Parameter	MU	MA1,1	MA1,2	MA2,1
MU	1.000	0.027	-0.014	-0.044
MA1,1	0.027	1.000	0.221	-0.225
MA1,2	-0.014	0.221	1.000	-0.123
MA2,1	-0.044	-0.225	-0.123	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	3.78	3	0.287	-0.004	0.013	0.018	0.161	0.028	0.120
12	13.81	9	0.129	0.005	-0.062	0.212	0.196	0.099	-0.079
18	19.21	15	0.204	0.131	0.040	0.058	-0.052	0.108	-0.120
24	24.58	21	0.266	0.159	-0.071	-0.082	-0.003	0.066	0.077

Table 7. Sample diagnostic statistics for the 1SW salmon model, 1971 to 1988.

Parameter	Estimate	Approx. Std Error	T Ratio	Lag
MU	0.05249	0.15492	0.34	0
AR1,1	0.18677	0.11120	1.68	2
AR1,2	0.09603	0.11068	0.87	3
AR1,3	0.27893	0.11240	2.48	4
AR2,1	-0.67521	0.08354	-8.08	5

Constant Estimate = 0.03853889
Variance Estimate = 1.18747215
Std Error Estimate = 1.08971196
AIC = 261.169519
SBC = 273.323603
Number of Residuals= 84

Correlations of the Estimates

Parameter	MU	AR1,1	AR1,2	AR1,3	AR2,1
MU	1.000	0.018	-0.001	-0.017	-0.000
AR1,1	0.018	1.000	-0.190	-0.251	-0.096
AR1,2	-0.001	-0.190	1.000	-0.204	-0.008
AR1,3	-0.017	-0.251	-0.204	1.000	-0.089
AR2,1	-0.000	-0.096	-0.008	-0.089	1.000

Autocorrelation Check of Residuals

To Lag	Chi Square	DF	Prob	Autocorrelations					
6	1.78	2	0.411	0.110	0.024	0.025	0.047	-0.055	0.040
12	4.85	8	0.774	0.032	0.062	0.007	-0.061	-0.059	-0.137
18	20.33	14	0.120	0.128	-0.069	-0.049	-0.263	-0.050	-0.223
24	22.09	20	0.336	-0.081	-0.054	0.040	0.010	0.015	-0.063

Table 8. Box-Jenkins forecast models of MSW salmon counts at Millbank trapnet.

	1971 to 1985 forecast 1986	1971 to 1986 forecast 1987	1971 to 1987 forecast 1988	1971 to 1988 forecast 1989
Parameters *				
Trend	-0.07123	-0.04245	-0.04542	-0.04093
RMA(1)	-0.38375	-0.40421	-0.41498	-0.29930
RMA(2)	-0.30906	-0.28315	-0.27378	-0.24629
SMA(5)	0.61096	-0.65242	0.67694	0.62549
Overall Mean	4.62	4.61	4.57	4.52
Residual Mean	0.092	0.093	0.093	0.086
Mean % Error	-5.25	-4.96	-4.57	-5.45
Absolute % Error	22.75	22.88	22.25	23.73
Index of Determination	0.17	0.14	0.13	0.07
Forecast ------------------ using models				
Period 1	95 (137)	163 (57)	145 (85)	94 ( )
Period 2	180 (102)	216 (66)	204 (57)	263 ( )
Period 3	211 (90)	155 (77)	150 (6)	154 ( )
Period 4	115 (28)	45 (33)	36 (146)	24 ( )
Period 5	37 (99)	44 (67)	52 (25)	46 ( )
Total	640 (456)	623 (300)	586 (319)	581 ( )
Confidence Int. 95%				
Period 1	2 - 557	5 - 930	5 - 797	2 - 541
Period 2	4 - 1064	6 - 1237	7 - 1124	7 - 1526
Period 3	5 - 1248	4 - 889	5 - 828	4 - 897
Period 4	3 - 682	1 - 256	1 - 199	1 - 141
Period 5	1 - 221	1 - 252	2 - 288	1 - 265
Using Overall Mean	507	502	483	459

* Trend equals mean of series after differencing
RMA( ) equals regular moving average term of lag ( )
SMA( ) equals seassonal moving average term of lag ( )

Table 9. Box-Jenkins forecast models of 1SW salmon counts at Millbank trapnet.

	1971 to 1985 forecast 1986	1971 to 1986 forecast 1987	1971 to 1987 forecast 1988	1971 to 1988 forecast 1989
Parameters *				
Trend	-0.07795	0.05222	0.07540	0.05249
RAR(1)	0.22468	0.21598	0.19096	0.18677
RAR(2)	0.14679	0.15470	0.13843	0.09603
RAR(3)	0.18812	0.23150	0.25597	0.27893
SAR(5)	-0.63391	-0.74278	-0.67657	-0.67521
Overall Mean	5.06	5.09	5.08	5.09
Residual Mean	0.001	-0.002	0.000	-0.002
Mean % Error	-69.28	-64.95	-62.97	-59.03
Absolute % Error	85.15	81.68	80.05	76.24
Index of Determination	0.69	0.64	0.63	0.60
Forecast				
using models				
Period 1	104 (124)	29 (32)	43 (60)	8 ( )
Period 2	3395 (746)	1566 (663)	2511 (619)	2087 ( )
Period 3	1910 (409)	1034 (366)	2068 (88)	776 ( )
Period 4	437 (95)	62 (39)	134 (905)	44 ( )
Period 5	142 (386)	122 (105)	412 (127)	75 ( )
Total	5989 (1760)	2812 (1205)	5168 (1799)	2990 ( )
Confidence Int. 95%				
Period 1	1 - 672	1 - 189	0 - 287	0 - 52
Period 2	36 -21971	12 -10384	19 -16664	16 -13905
Period 3	20 -12415	8 - 6875	15 -13761	6 - 5167
Period 4	4 - 2850	0 - 413	1 - 893	0 - 296
Period 5	1 - 931	1 - 814	3 - 2756	1 - 503
Using Overall Mean	788	812	804	812

* Trend equals mean of series after differencing
RAR( ) equals regular auto-regressive term of lag ( )
SAR( ) equals seasonal auto-regressive term of lag ( )

Table 10. Summary of residuals and number of years when forecast was >±30% of returns using Gaussian and Cauchy estimators and the likelihood and least squares approaches to smoothing parameter selection. Single, refers to a single smoothing parameter. Ind. refers to a model in which a smoothing parameter was calculated for each variable in the model.

Model	Residual	Number of years >30%
1971-1989		
Mean	2064	8
Likelihood		
Raw Normal (Single)	1601	7
Raw Normal (Ind)	1714	6
Ln Normal (Single)	1623	5
Ln Normal (Ind)	1649	7
Least Squares		
Raw Normal (Single)	1562	7
Raw Normal (Ind)	1551	7
Ln Normal (Single)	1623	5
Ln Normal (Ind)	1623	5
Cauchy	1523	6
1974-1989		
Mean	1730	7
Cauchy	1250	5
Cauchy (GREEN)	970	5
Cauchy (SFA4)	1191	5

Table 11. Smoothing parameters chosen for Gaussian and Cauchy models using maximum likelihood and least squares validation approaches.

Procedure	Raw-normal Individual 1SW	MSW	Single	Ln-normal Individual 1SW	MSW	Single
Likelihood	0.30	0.85	0.65	0.45	0.85	0.70
Least Sq.	0.55	0.50	0.50	0.70	0.70	0.70
Cauchy (Raw MSW)		17				
Cauchy (GREEN)		10,000				
Cauchy (SFA4)		50				

Table 12. Residuals (forecast-returns) for Cauchy distribution model using 1SW year i-1 to forecast MSW year i returns from 1972 to 1990.

MSW Year	Residual
72	-5
73	2
74	-16
75	-1
76	6
77	-24
78	3
79	15
80	-9
81	12
82	-3
83	0
84	9
85	3
86	-3
87	3
88	6
89	5
90	0

Table 13. Inseason forecasts of MSW salmon returns to the Miramichi River for standardized weeks 22 and 25. Inseason total forecast is the sum of cumulative return to date + MSW salmon to return forecast. Total forecast is (w x inseason total forecast) + ((1-w) x preseason forecast).

Year	MSW Salmon to Return Forecast	Inseason Total Forecast	Inseason 95% C.I. Width (B)	Preseason 95% C.I. Width (A)	Weight (w)	Total Forecast	Resdiual
Standardized Week 22							
1977	20	22	46	25	0.352	26	-26
1979	20	21	45	31	0.408	20	-11
1983	22	27	29	40	0.580	26	-2
1986	21	23	45	34	0.430	28	-3
Standardized Week 25							
1977	15	21	43	25	0.368	25	-27
1979	14	16	42	31	0.425	16	-7
1983	11	29	22	40	0.645	27	-1
1986	17	30	30	34	0.531	31	0

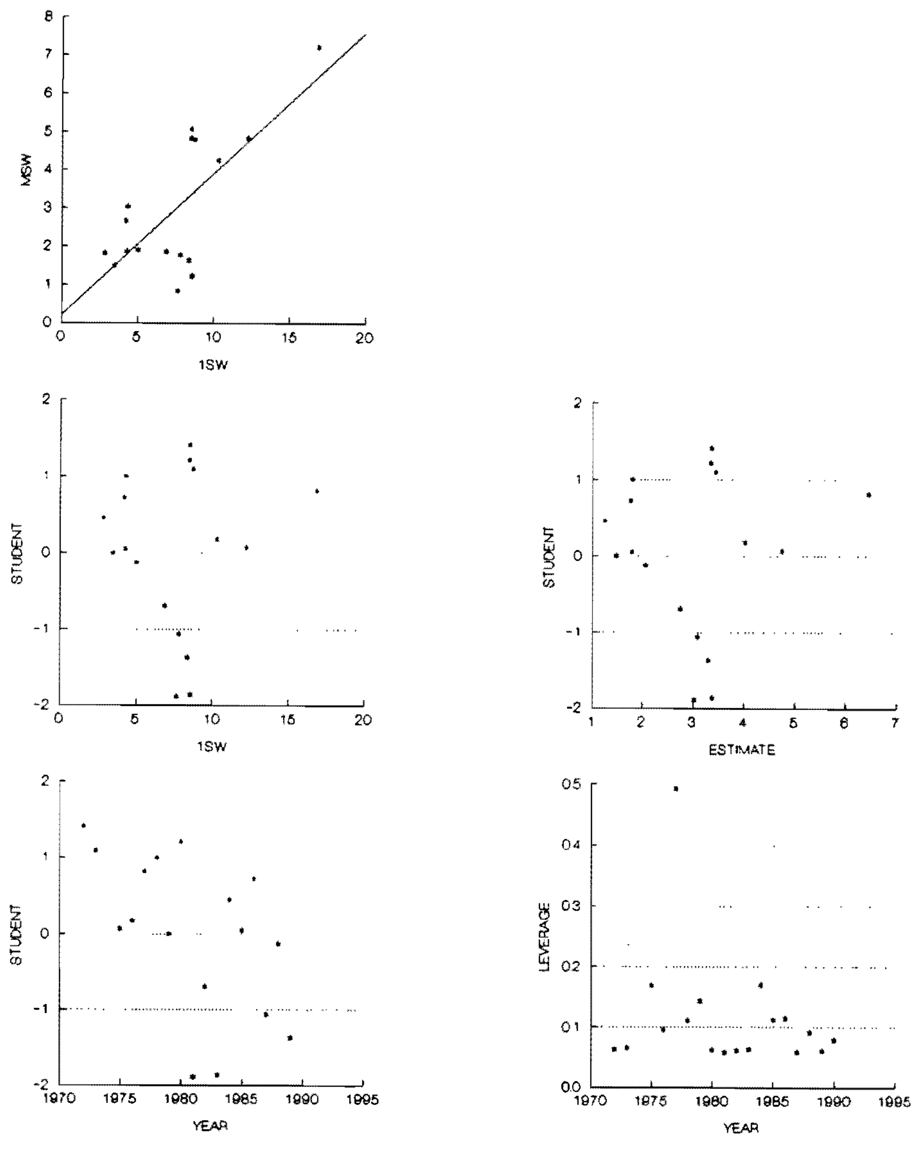


Figure 1. Correlation between abundance of 1Sw salmon (year i-1, expressed as CPUE) and MSW salmon (year i) at Millbank, 1971 to 1989 (upper left figure). See regression equation (2) in text. Residual and leverage plots are also shown (lower 4 figures).

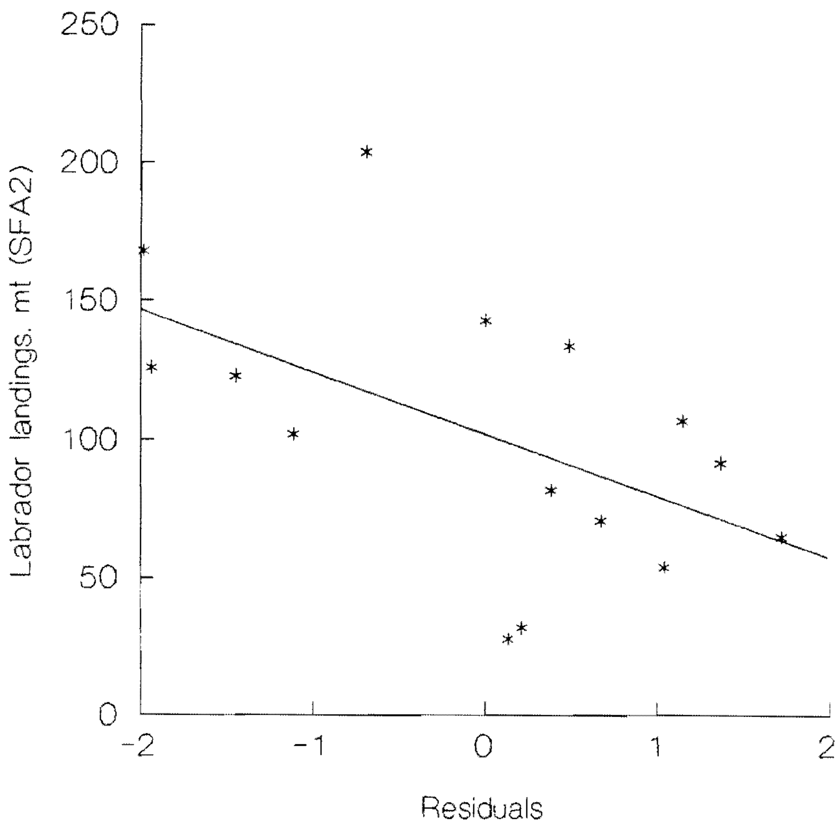


Figure 2. Relationship between residuals from the 1SW to MSW salmon regression (equation (2) in text) and landings of small salmon in SFA 2 (Labrador). The regression coefficient was significant at $P<0.05$.

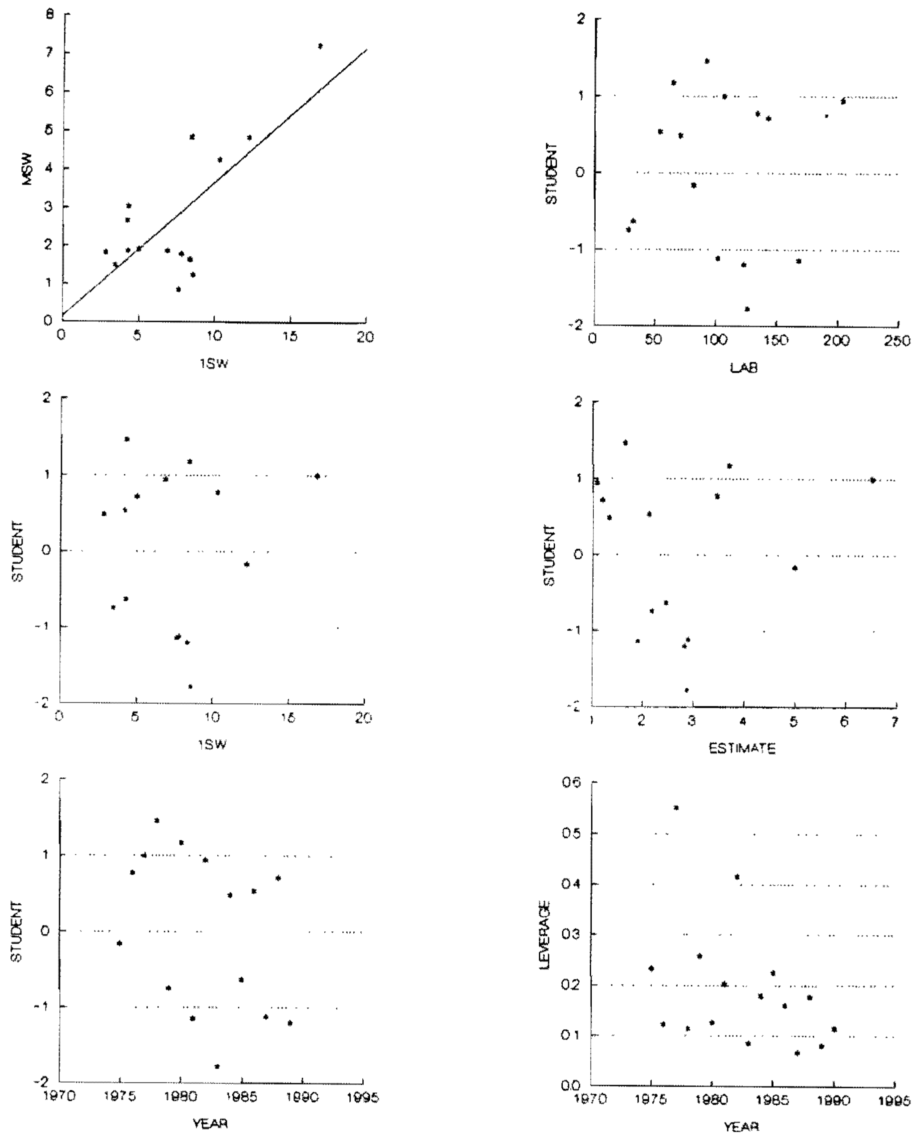


Figure 3. Residual and leverage plots for the multiple regression equation (3) (see text) with MSW salmon as the dependent variable and 1SW salmon and Labrador landings (LAB) as the independent variables.

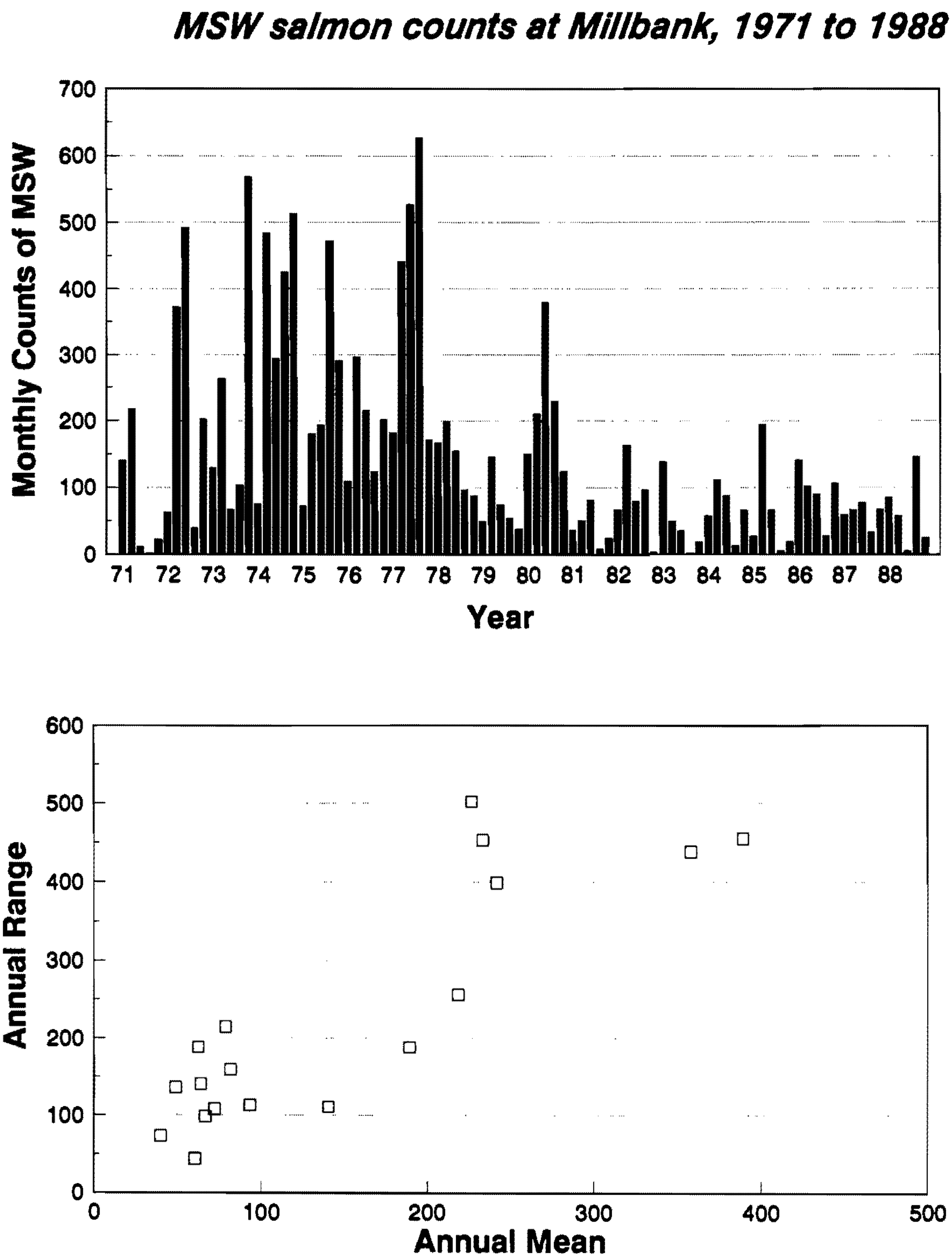
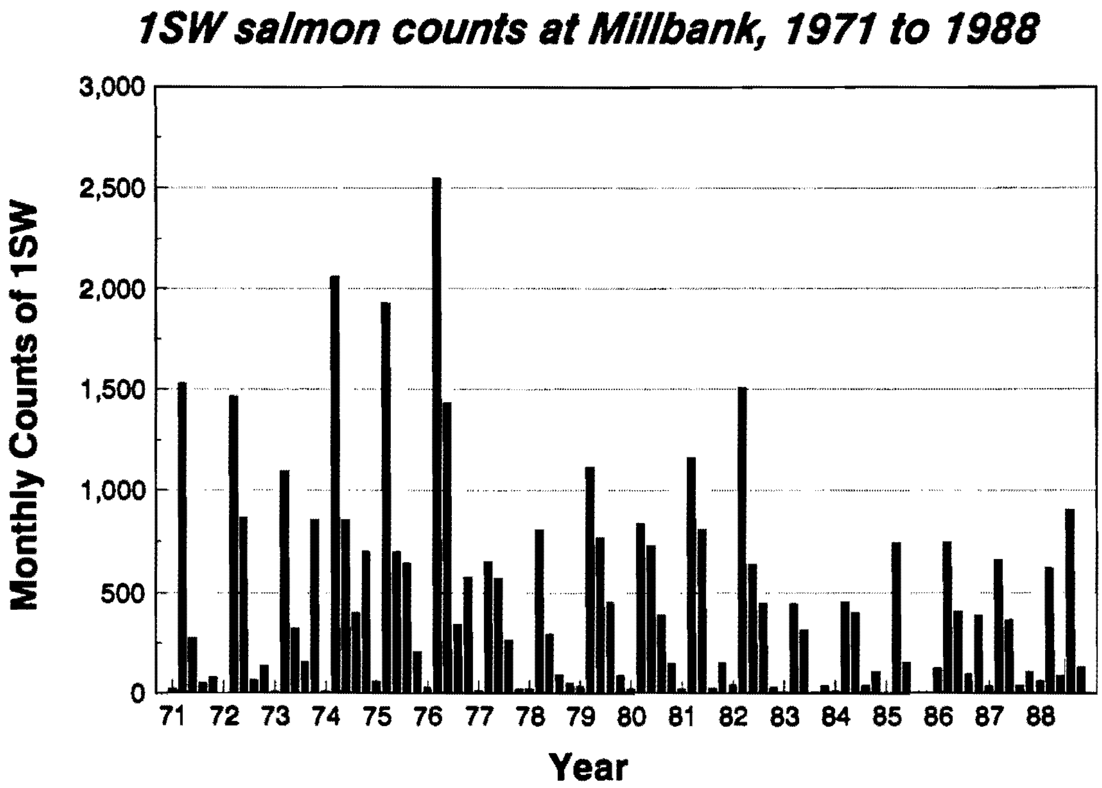


Figure 4. Counts of MSW salmon at Millbank, 1971-1988, for each monthly period and ranges versus mean scatterplot for each year.



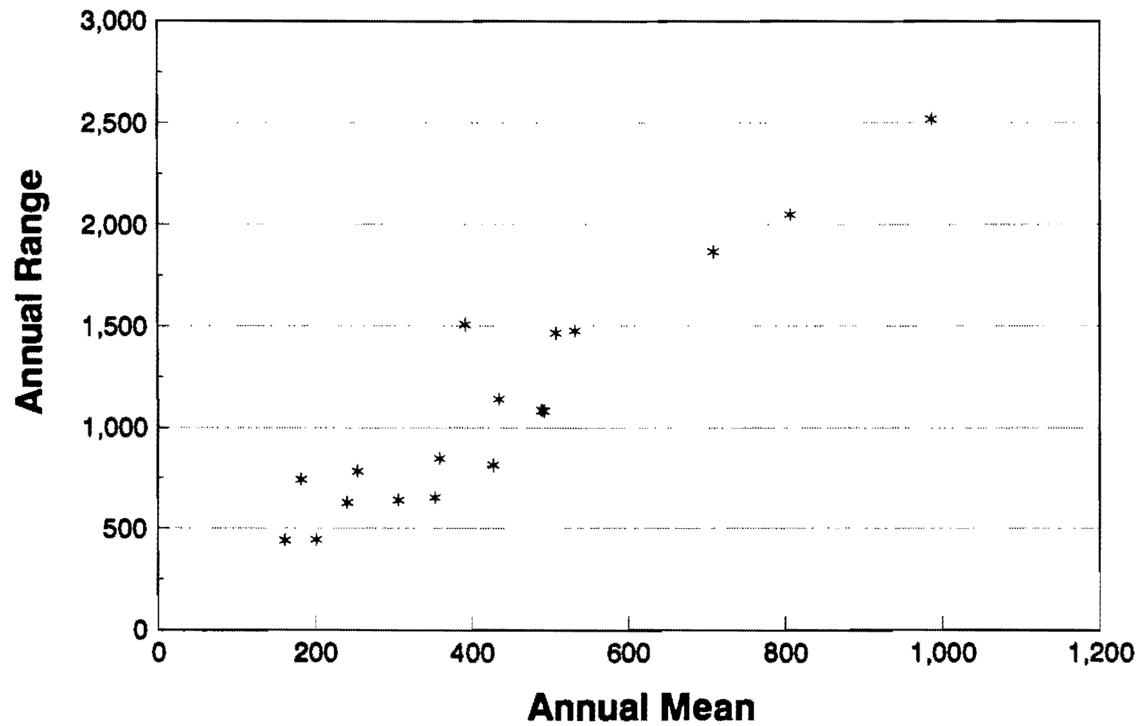


Figure 5. Counts of 1SW salmon at Millbank, 1971-1988, for each monthly period and ranges versus mean scatterplot for each year.

Fig. 6. Autocorrelation plot of the residuals of the MSW salmon model, 1971 to 1988.

```
Lag Covariance Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
-----------------------------------------------------------------------
  1 -0.0048463    -0.00408  |                .   |   .                |
  2   0.015449     0.01300  |                .   |   .                |
  3   0.021467     0.01807  |                .   |   .                |
  4   0.190993     0.16077  |                .   |***.                |
  5   0.033634     0.02831  |                .   |*  .                |
  6   0.141986     0.11952  |                .   |** .                |
  7  0.0064461     0.00543  |               .    |    .               |
  8  -0.073448    -0.06183  |               .   *|    .               |
  9   0.252170     0.21227  |               .    |****.               |
 10   0.232469     0.19568  |               .    |****.               |
 11   0.117118     0.09859  |               .    |**  .               |
 12  -0.093787    -0.07895  |               .  **|    .               |
 13   0.155870     0.13121  |               .    |*** .               |
 14   0.047596     0.04006  |               .    |*   .               |
 15   0.068402     0.05758  |               .    |*   .               |
 16  -0.062231    -0.05238  |               .   *|    .               |
 17   0.128761     0.10839  |               .    |**  .               |
 18  -0.142750    -0.12016  |               .  **|    .               |
 19   0.188732     0.15887  |               .    |*** .               |
 20  -0.084538    -0.07116  |               .   *|    .               |
 21  -0.097600    -0.08216  |               .  **|    .               |
 22 -0.0035250    -0.00297  |               .    |    .               |
 23   0.078427     0.06602  |               .    |*   .               |
 24   0.091008     0.07661  |               .    |**  .               |
                             "." marks two standard errors
```

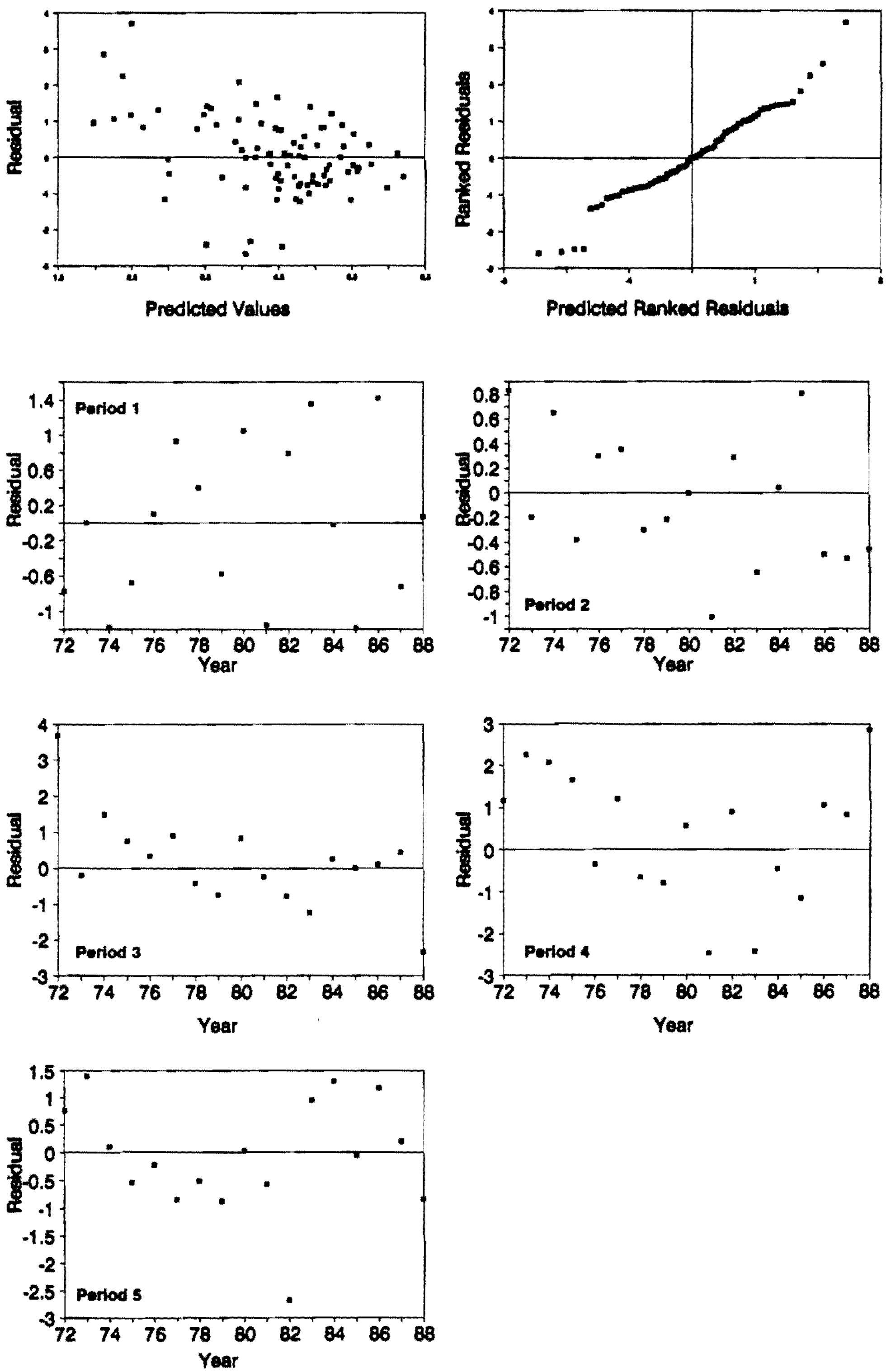


Figure 7. Residual analysis of MSW salmon counts at Millbank, 1971-1988, for five time periods.

Fig. 8. Autocorrelation plot of the residuals of the 1SW salmon model, 1971 to 1988.

```
Lag Covariance Correlation -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
------------------------------------------------------------------------
  1   0.130362     0.10978 |                .   |** .                |
  2   0.028816     0.02427 |                .   |   .                |
  3   0.029911     0.02519 |                .   |*  .                |
  4   0.055223     0.04650 |                .   |*  .                |
  5  -0.064846    -0.05461 |                .  *|   .                |
  6   0.047900     0.04034 |                .   |*  .                |
  7   0.038359     0.03230 |                .   |*  .                |
  8   0.073829     0.06217 |                .   |*  .                |
  9  0.0077920     0.00656 |                .   |   .                |
 10  -0.072320    -0.06090 |                .  *|   .                |
 11  -0.070610    -0.05946 |                .  *|   .                |
 12  -0.163021    -0.13728 |               . ***|    .               |
 13   0.151794     0.12783 |               .    |*** .               |
 14  -0.081539    -0.06867 |               .   *|    .               |
 15  -0.058247    -0.04905 |               .   *|    .               |
 16  -0.312667    -0.26330 |               *****|    .               |
 17  -0.059135    -0.04980 |               .   *|    .               |
 18  -0.264960    -0.22313 |               .****|    .               |
 19  -0.096702    -0.08144 |               .  **|    .               |
 20  -0.064309    -0.05416 |               .   *|    .               |
 21   0.047110     0.03967 |               .    |*   .               |
 22   0.011412     0.00961 |               .    |    .               |
 23   0.018231     0.01535 |               .    |    .               |
 24  -0.074389    -0.06264 |               .   *|    .               |
                           "." marks two standard errors
```

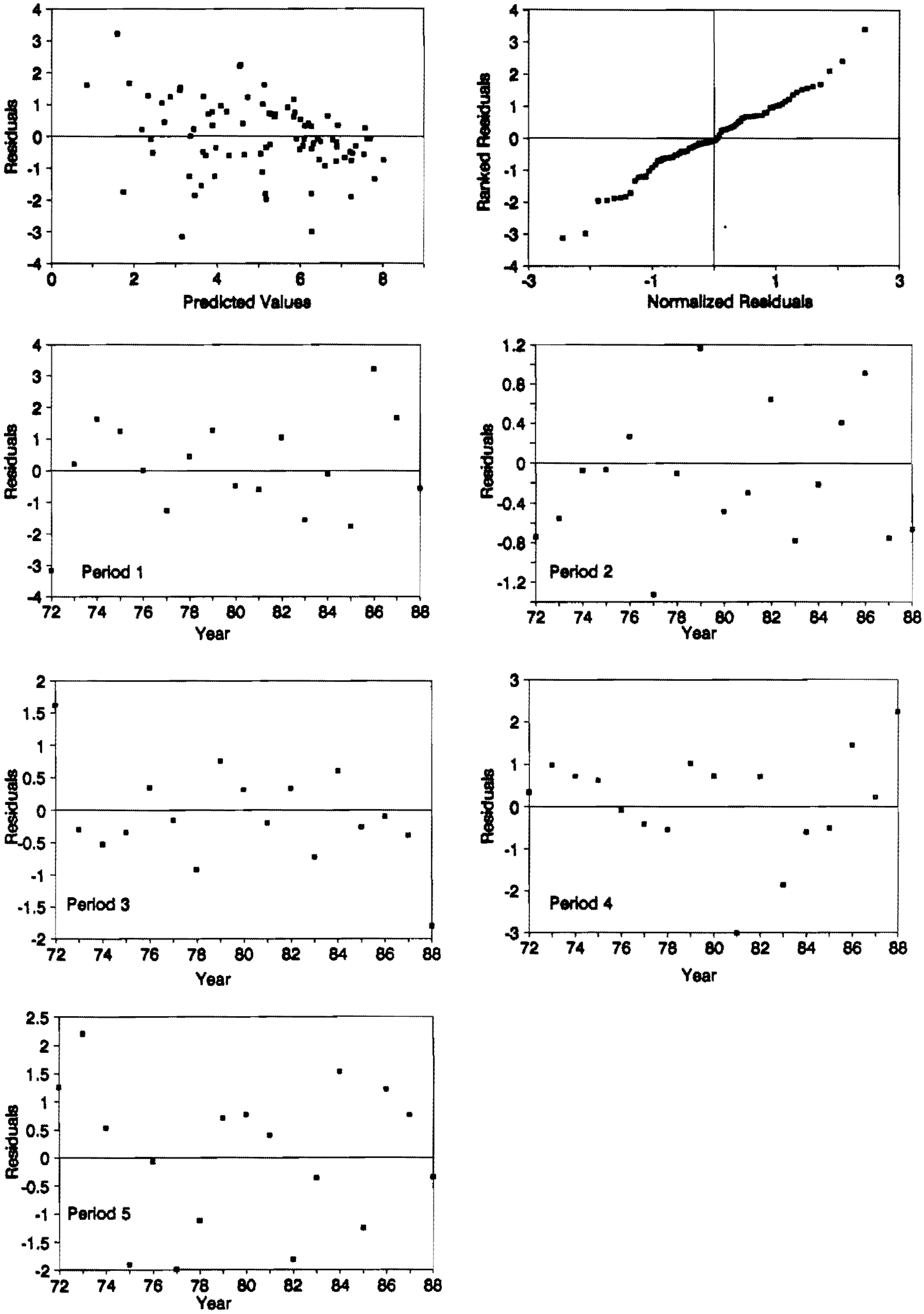
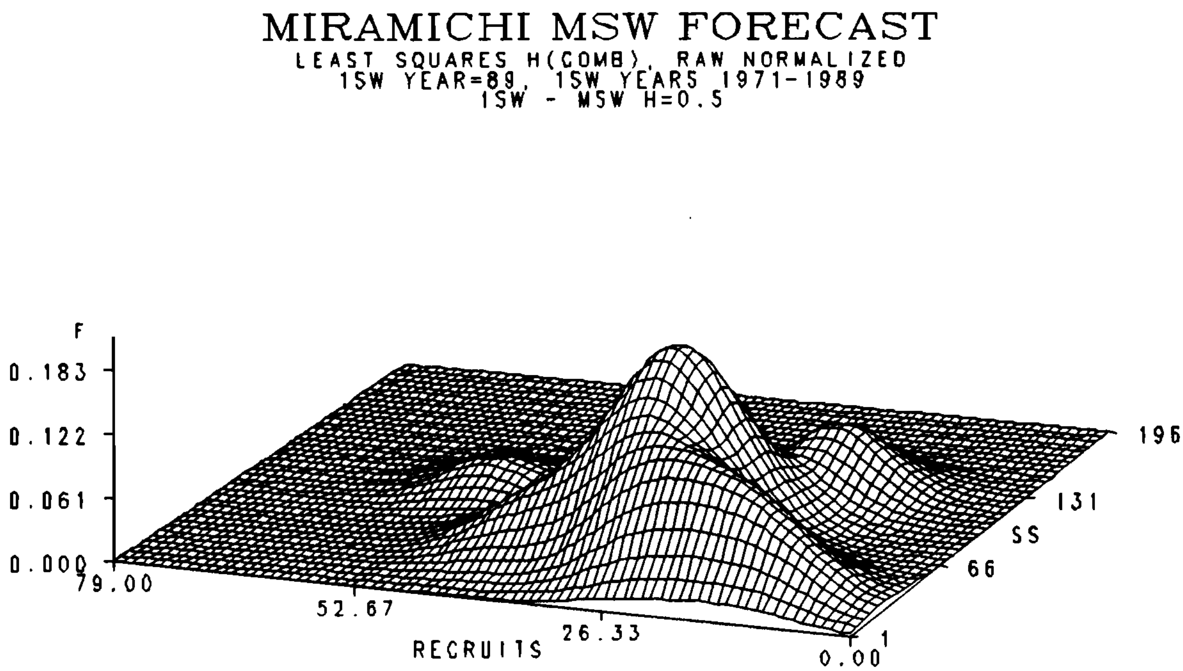


Figure 9. Residual analysis for 1SW salmon counts at Millbank from 1971-1988.



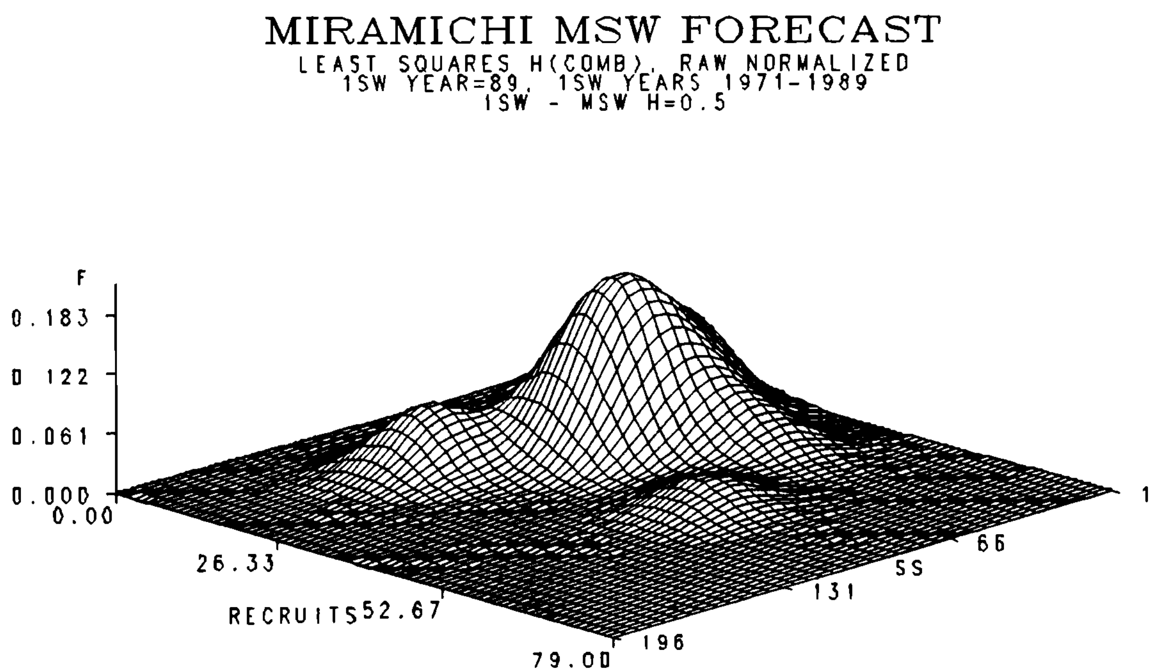
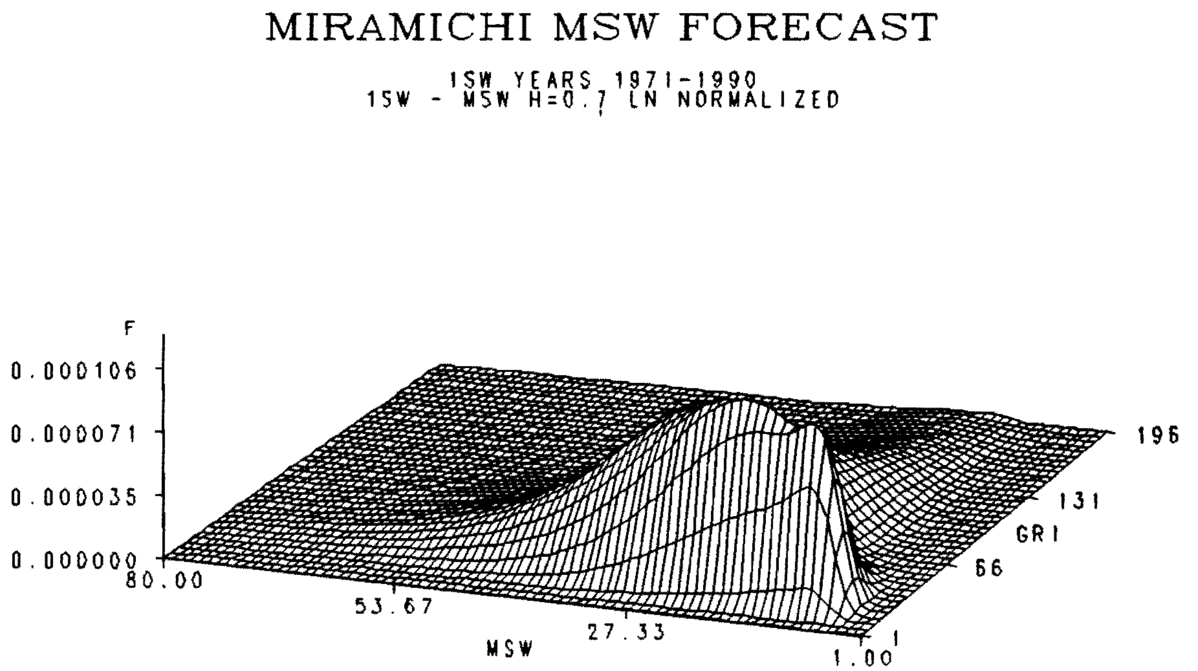


Figure 10. Gaussian distribution for Miramichi salmon based on raw normalized data with least squares smoothing parameter selection. 1SW (SS) and MSW (recruits).



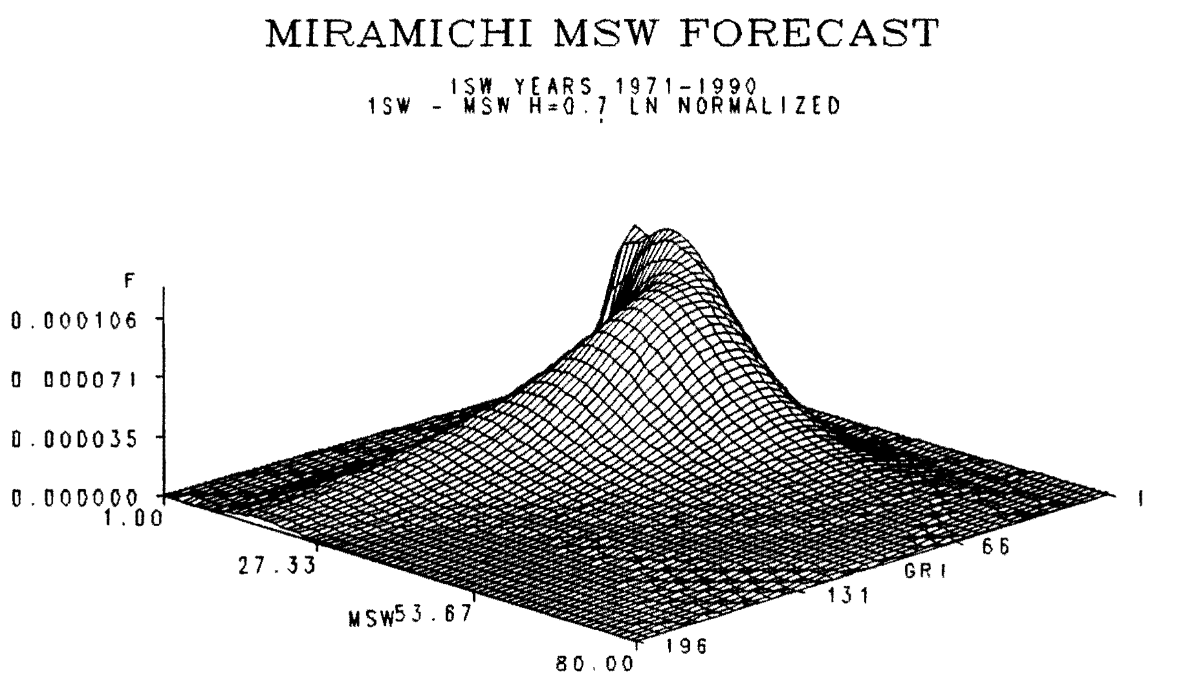


Figure 11. Gaussian distribution for Miramichi salmon using transformed logged data with maximum likelihood smoothing parameter selection. 1SW (SS) and MSW (recruits).

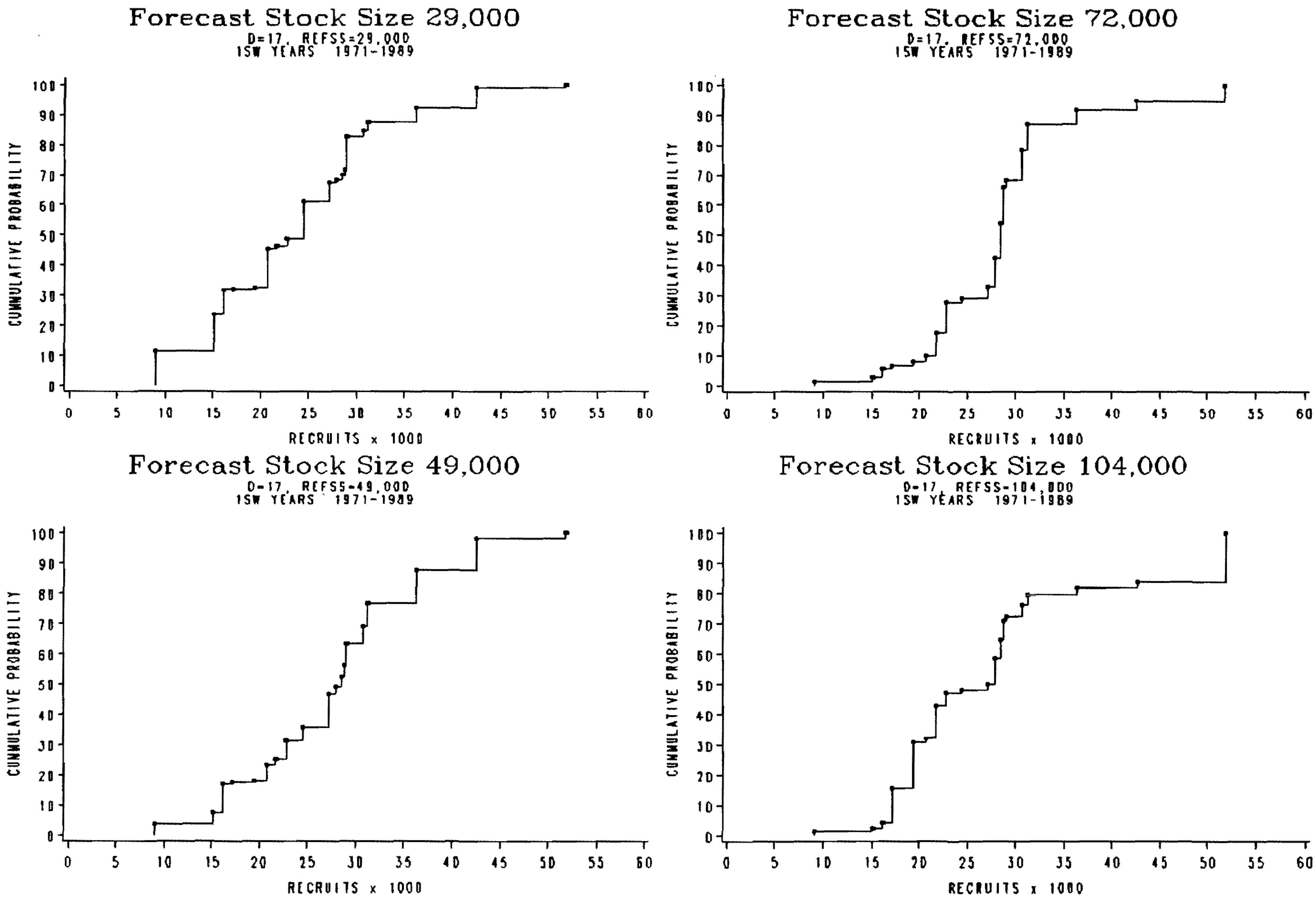


Figure 12. Cummulative frequency distributions at four stock (1SW) levels based on Cauchy distribution for Miramichi salmon.

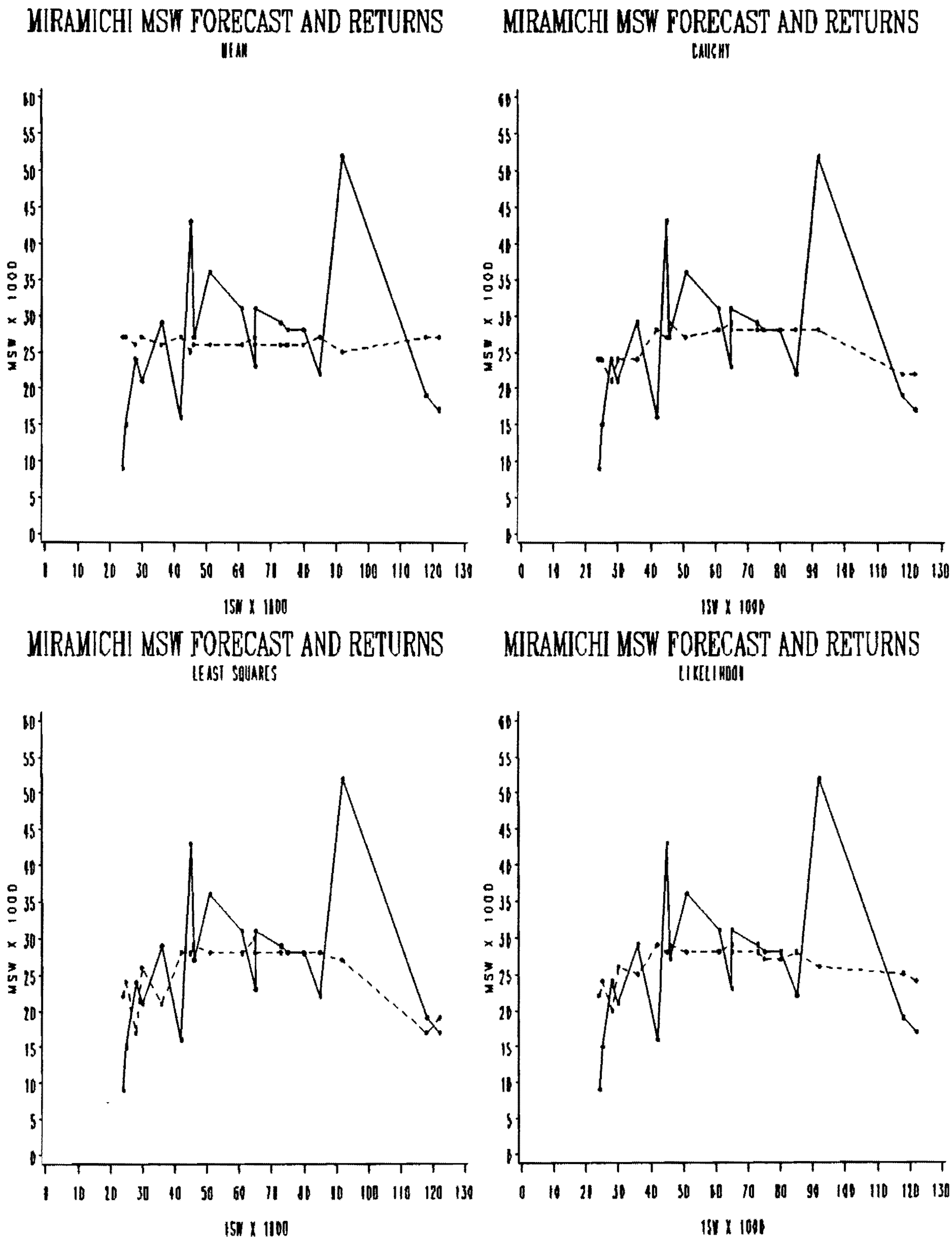


Figure 13. MSW forecasts (dashed line) and returns (solid line) for mean and three probability distribution models for Miramichi salmon.

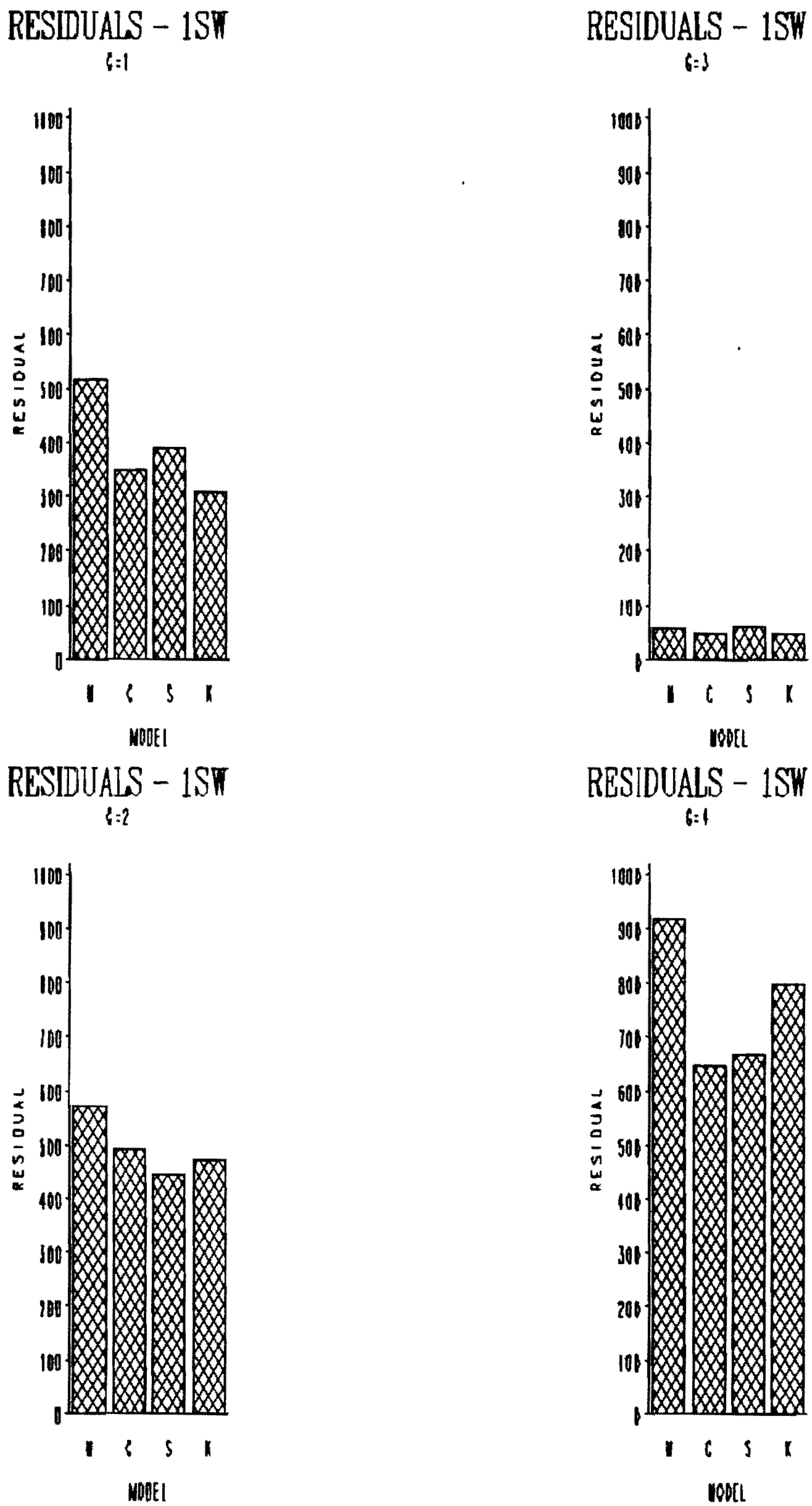


Figure 14. Residual sum of squares for mean (M), Cauchy (C), raw normalized least square (S), and ln normalized likelihood (K) models at low (G=1), intermediate (G=2, G=3) and high (G=4) stock levels for Miramichi salmon.

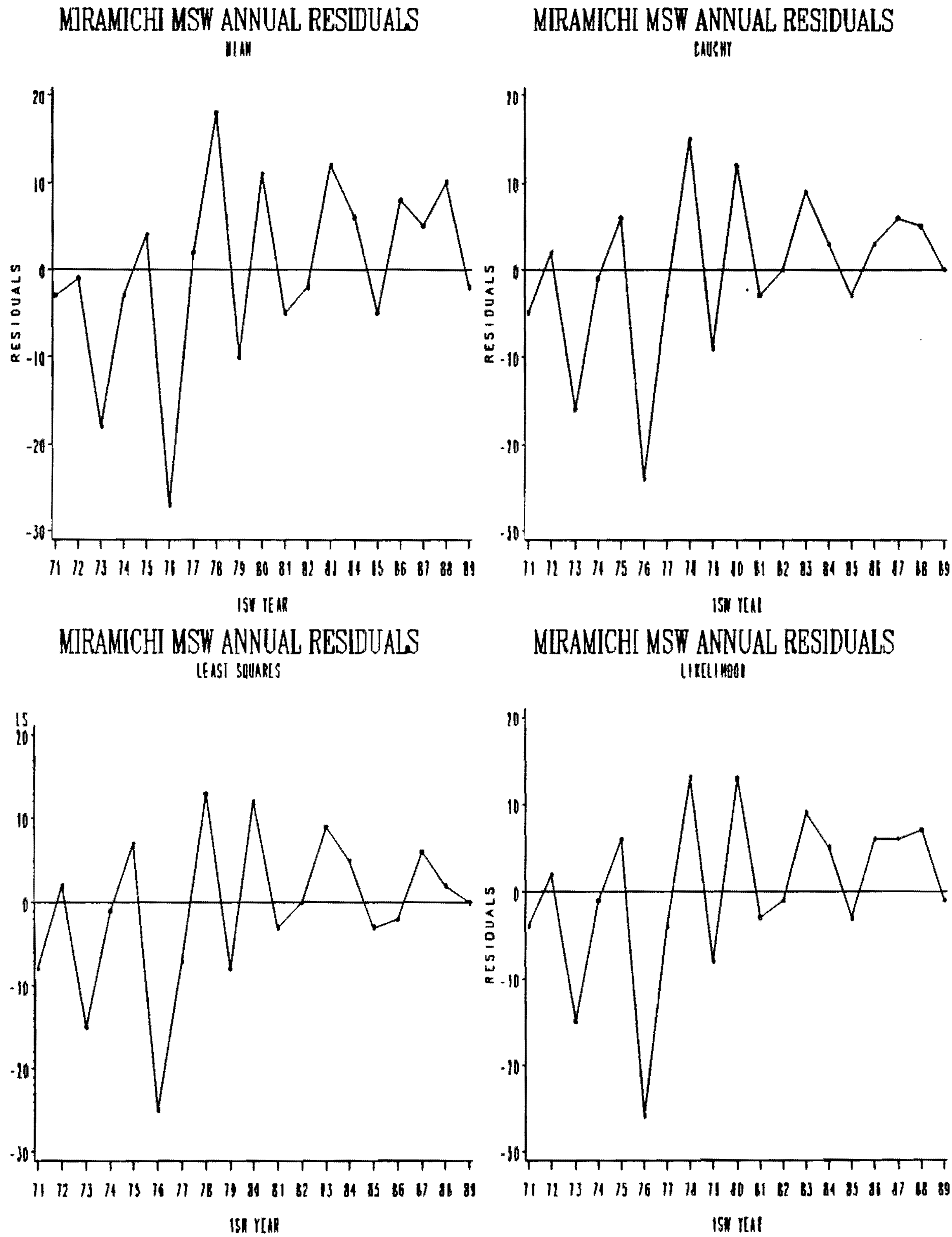


Figure 15. Annual residual trends for mean and three probability models.

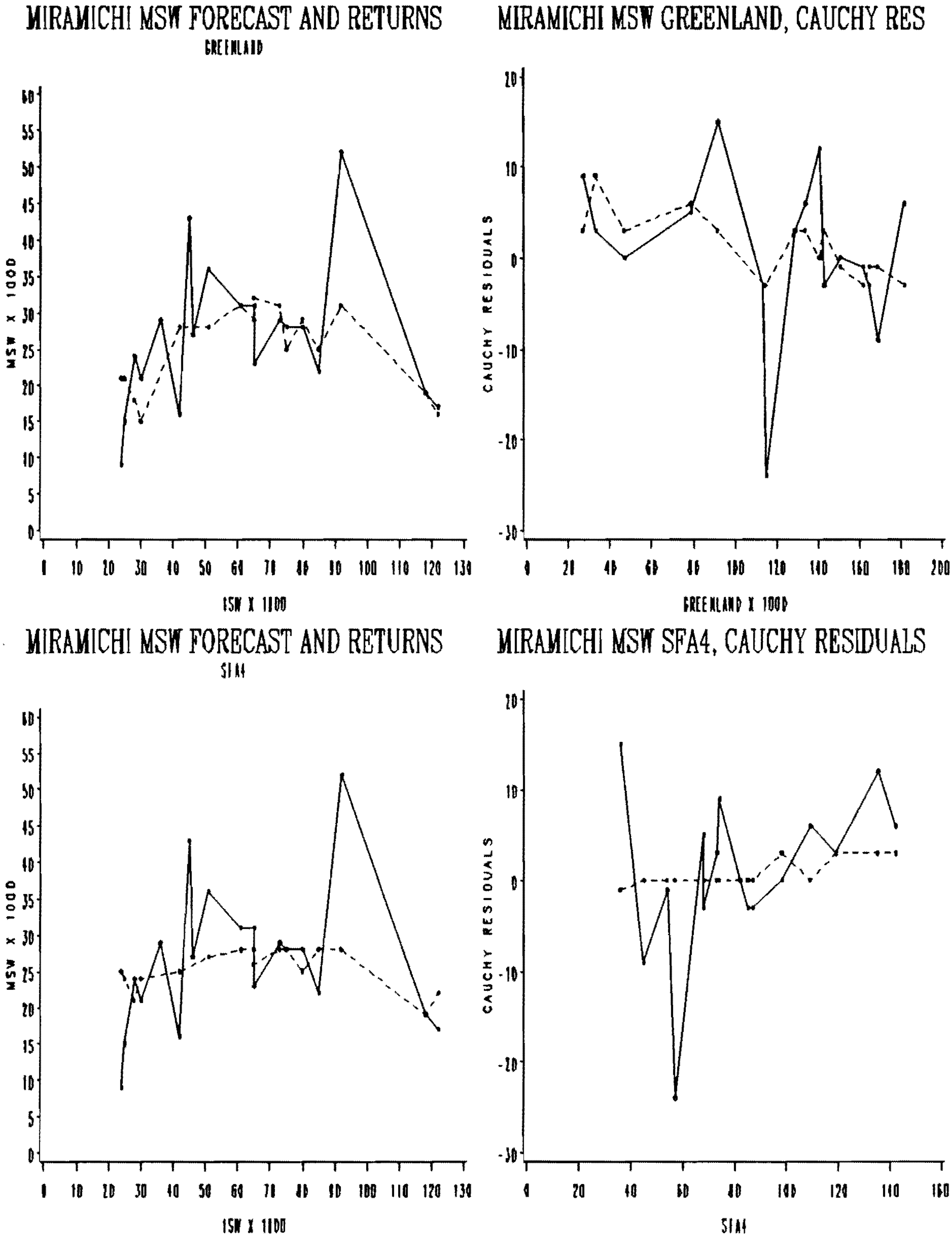


Figure 16. MSW forecasts and returns using GREEN and SFA4 data sets plotted against 1SW stock levels (left side) and forecast residuals (dashed line) using GREEN and SFA4 data sets (dashed line) and 1SW forecast residuals (solid line) (right side).

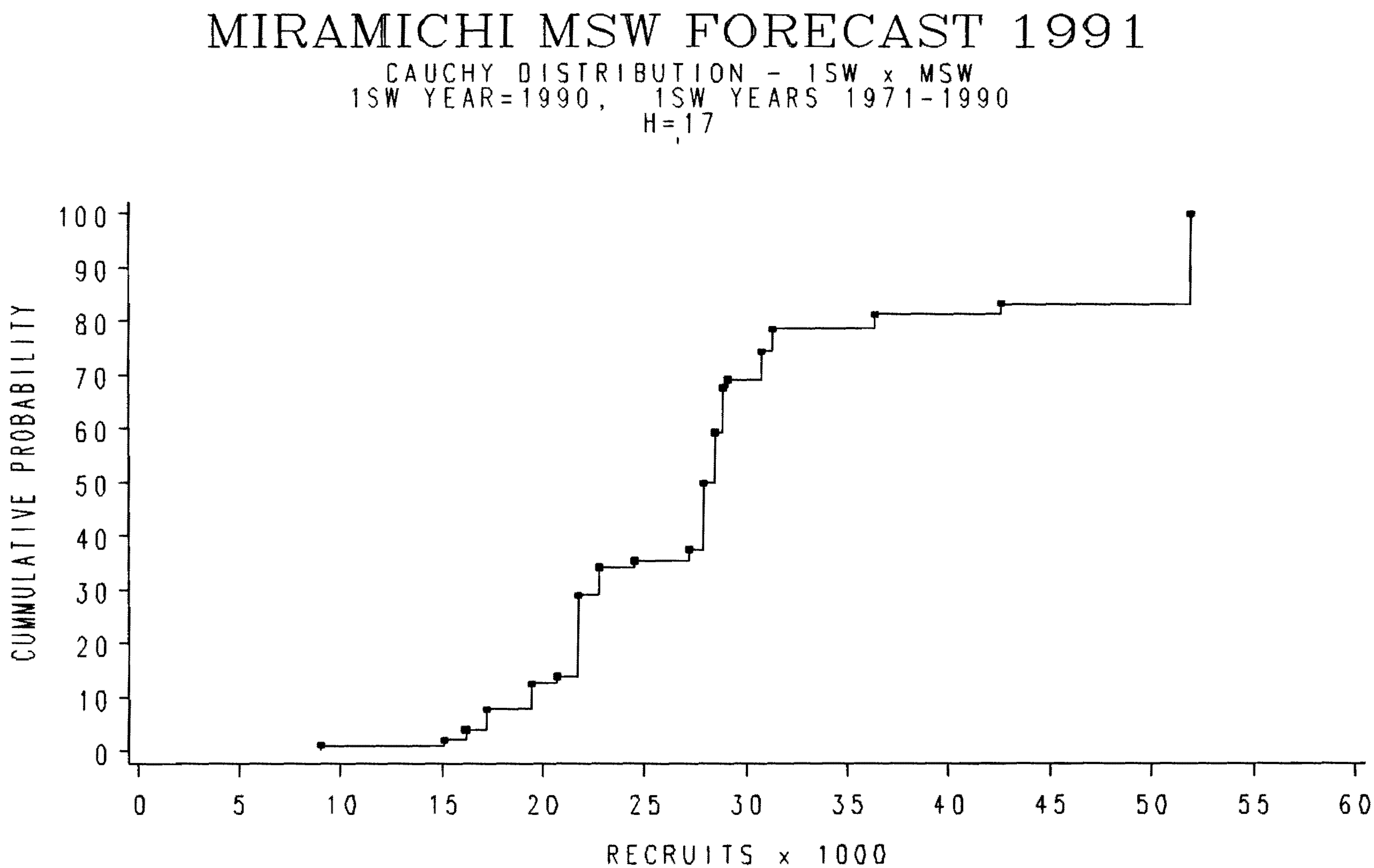


Figure 17. Cummulative frequency for Cauchy model for 1991 MSW Miramichi salmon forecast.

Appendix 1. Programs to calculate smoothing parameters and probability distributions used in preseason forecasts. The programs used for the time series analysis require run to date and run yet to return data as inputs for stock and recruitment.

A. Estimates smoothing parameter for Cauchy kernel estimator.

```
* k1.sas;
************************************;
title 'MIRAMICHI MSW FORECAST';
TITLE2 'MSW H, SFA4 - CAUCHY';
TITLE3 '1SW YEARS 1974-1989';
************************************;
* Estimates kernel estimater for pdf forecast model
    using cross-validation leave-one-out procedure.
  See Evans and Rice. 1988. Predicting recruitment
    from stock size ... J. Cons. Explor. Mer. 44: 111-122.
    Rice and Evans. 1988.  Tools for embracing uncertainty
    in the management of the cod .. J. Cons. Explor. Mer. 45:73-81. For details regarding methodology.;
 ********* Value for D,  Data step RC2, is determined by iteration.
    A lower value for D is chosen in each run until the variance in
      50% recruitment is minimized.  Variance in this case is the       sum of squared residuals.  The last value for NEWVAR
in data        set RC3.;
*************** ERASE OLD K.OUT FILE BEFORE NEW RUNS;
 ********* New value for d, data step RC2  must be entered
     in each run;
**************** The first run through use PROC MEANS to get summary
     statistics for data set.  Comment it out for subsequent runs.;
*options linesize=132;

data a(KEEP=GRYEAR YEAR SS RECRUITS);
***************************;
****************************;
infile 'dua1:[claytor.forecast]RESMIR.dat' missover;
*****************************;
*****************************;
*INPUT FOR RESMIR.DAT;
 INPUT GRYEAR GRI MSW CAUCHY LS LK MEAN GREEN SFA4;
 IF GRYEAR LE 73 THEN DELETE;
 SS=SFA4;
 RECRUITS=CAUCHY;
 YEAR=GRYEAR;
RUN;
 * input for mir.dat;
 * input year ss recgr recsal;
 * recruits=sum(recgr,recsal);
 /*
 *INPUT FOR MIR1SW.DAT;
 * INPUT GRYEAR GR SAL sfa2 sfa3 sfa4 RESIDUAL;
* input for miram.dat;
    input gryear gr sal sfa2 sfa3 sfa4 residual sfa5 sfa6 sfa7
          sfa8 sfa10 sfa12 res86/ res87 RES88 RES89;
  SS=GR/1000;
  RECRUITS=SAL/1000;
  *SS=SFA4;
  *RECRUITS=RESIDUAL;
  YEAR=GRYEAR;
 * TO MAKE DATA CONSISTENT WITH NEWFOUNDLAND CATCH DATA
     DELETE ALL YEARS UP TO 1974;
 *IF YEAR LE 73 THEN DELETE;
 *DELETE YEAR-1 TRYINGO PREDICT, I.E. for MSW 1989
   DELETE YEAR 88 GE GRYR 88
     if using all years up to year trying to predict.
     i.e. 1971-1987  THEN MUST USE GE;
  ****************************;
 * if year ge 83 then delete;
******************************;
 *MUST DELETE YEAR=90 BECAUSE NO RECRUITS I.E. NO 1990 RECRUITS
    YET;
  IF YEAR=90 THEN DELETE;
```

```
*/
/*
 *INPUT FOR MIRCOUNT.DAT;
 INPUT GRYEAR E1SW EMSW L1SW LMSW LAB;
  T1SW=SUM(E1SW,L1SW);
  TMSW=SUM(EMSW,LMSW);
  SS=T1SW/100;
  RECRUITS=TMSW;
  YEAR=GRYEAR;
   IF YEAR=80 OR YEAR=84 THEN DELETE;
  *DELETE YEAR TRYING TO PREDICT, I.E. 1989 MSW COUNTS;
   IF YEAR=88 OR YEAR=89 THEN DELETE;
 */

 *input for jake.dat;
 * input year ss recruits;

 *input for dempenv.dat;
*  input year ss recruits;
  * year=label ss=temp recruits=smolt count;
run;
/*
DATA A;
  INFILE 'DUA1:[CLAYTOR.FORECAST]FAKE.DAT' MISSOVER;
  INPUT GRYEAR GRI SFA4 MSW;
  YEAR=GRYEAR;
  SS=GRI;
  RECRUITS=MSW;
  N=4;
RUN;

PROC MEANS DATA=A N MEAN STD MIN MAX;
  ID N;
  VAR GRI SFA4 MSW;
  OUTPUT OUT=B
      N=NGRI
      MEAN=MEANGRI MEANSFA4 MEANMSW
      STD=STDGRI STDSFA4 STDMSW;
RUN;
PROC PRINT DATA=A; TITLE5 'FAKE.DAT'; RUN;
PROC PRINT DATA=B;  TITLE5 'DATA=B MEANS'; RUN;
*/
proc print data=a;
  TITLE5 'Miramichi total returns mir1sw.dat';
  TITLE5 'data=a original data set';
run;

proc means data=a;
  var ss recruits;
run;

proc sort data=a out=rc;
  by recruits;
run;

proc print data=rc;
  TITLE5 'data=rc sorted data set';
 run;
proc sort data=a out=ss;
  by ss;
run;
proc print data=ss;
  title 'data=ss';
run;

****************;
****************;
data rc2(DROP=SUMF) totf(drop=recruits f);
  set rc;
```

```
   REFYR=year;
   REFSS=ss;
   REFREC=recruits;

**************;
  D=55;  *KERNEL ESTIMATER;
**************;

cumf=0;
* relate each observation to all the others;
do i=1 to count;
  set rc point=i nobs=count;
  if i=_n_ then go to nexti;
  x=REFSS-ss; f=1/ (1 + (x/d)**2 );
  cumf+f;
  output rc2;
nexti: end;
sumf=cumf; output totf;
run;
/*
proc print data=rc2; TITLE5 'data set rc2'; run;

proc print data=totf; TITLE5 'data set totf'; run;
*/
proc sort data=rc2; by REFREC refyr; run;
proc sort data=totf; by REFREC refyr; run;


data rcx3; merge rc2 totf(drop=refyr); by REFREC; run;

proc sort data=rcx3 out=srcx3;
  by refyr;
run;
data rc3;
  set srcx3;
   by refyr;
 retain flag;
 if first.REFyr then DO;
   flag=0;
   CPCT=0;
  END;
FPCT=100*F/SUMF;
CPCT+FPCT;
if flag=0 and CPCT ge 50 then do;
  flag=1;
  y=(REFREC-recruits)**2;
  NEWVAR+y;
*  output;  * OUTPUT IN ALL CASES;
end;
run;
/*
 proc print; TITLE5 'data set rc3'; run;
*/
OPTIONS LINESIZE=79;
data look(keep=d YEAR SS RECRUITS REFYR REFSS REFREC NEWVAR oldvar                  per);
  set rc3;
  if y=. then delete;
/*
*******************;
  OLDVAR=100;
********************;

  vardif=NEWVAR-oldvar;
  per=(vardif/oldvar)*100;
*/
run;

proc print data=look;
  TITLE5 'data=look variance estimates for each stock level';
```

```
run;

* find last line of  look.ssd  and append it to a summary data set;
/*
FILENAME OUTFIL 'DUA1:[CLAYTOR.FORECAST]KERNEL.OUT';
data _NULL_; set look nobs=count; file OUTFIL MOD;
if _n_=count then put D 12.6 NEWVAR 20.;
run;
*/
FILENAME OUTFIL 'DUA1:[CLAYTOR.FORECAST]K1.OUT';
  DATA _NULL_; SET LOOK NOBS=COUNT; FILE OUTFIL MOD;
  IF _N_=COUNT THEN PUT D 12.6 NEWVAR 20.;
RUN;

DATA X;
 INFILE 'DUA1:[CLAYTOR.FORECAST]K1.OUT' MISSOVER;
  input d variance;
RUN;
PROC SORT DATA=X OUT=SX;
  BY  DESCENDING D;
RUN;

PROC PRINT DATA=sx;
  var d variance;
  FORMAT D 9.2;
  TITLE5 'DATA=SX SUMMARY OF KERNEL ITERATIONS';
RUN;
  data _null_; set sx; file 'k.out';
  put d 12.6 variance 20.;
 run;
```

B. Program for estimating smoothing parameter for Gaussian kernel estimator using the least square validation procedure.

```
*LSHc.SAS;
****************************;
TITLE 'LEAST SQUARE VALIDATION SMOOTHING PARAMETER';
TITLE2 'MULTIVARIATE SINGLE H TWO VARIABLES IN MODEL';
*****************************;
/*
   CALCULATES SMOOTHING PARAMETER (H) USING LEAST
    SQUARES VALIDATION APPROACH.  FINDS H WITH THE
    LEAST SUM OF SQUARES ERROR IN RECRUITMENT
    FORECASTS. ANALOGOUS TO EVANS AND RICE APPROACH
    FOR CAUCHY ALGORITHM EXCEPT IT IS APPLIED TO
    GAUSSIAN KERNEL.
   ITEMS TO ENTER BY HAND ARE IN DATA STEP V, DO I, AND
    DO J. AND H OR SMOOTHING PARAMETER VECTOR ON THE
    LINE AFTER PROC IML.  THESE ARE MARKED BY A
    DOUBLE LINE OF ********.
*/

DATA A(KEEP=N SS REC REFYR YEAR);
  INFILE 'DUA1:[CLAYTOR.FORECAST]MIRAM.DAT' MISSOVER;
  INPUT GRYEAR GR SAL SFA2 SFA3 SFA4 RESIDUAL SFA5 SFA6 SFA7
        SFA8 SFA10 SFA12 RES86/RES87 RES88 RES89;
  GRI=GR/1000;
  MSW=SAL/1000;
  ******************;
  IF GRYEAR=90 THEN DELETE;
**********************;
  REFYR=GRYEAR;
  YEAR=GRYEAR;
  SS=GRI;
  REC=MSW;
  N=19;
RUN;
/*
DATA A;
  INFILE 'DUA1:[CLAYTOR.FORECAST]FAKE.DAT' MISSOVER;
  INPUT GRYEAR GRI SFA4 MSW;
  REFYR=GRYEAR;
  YEAR=GRYEAR;
  N=4;
  SS=GRI;
  REC=MSW;
RUN;
*/
DATA NA;
  SET A;
  IF YEAR=71 THEN DELETE;
RUN;
PROC MEANS DATA=NA N MEAN STD MIN MAX;
  ID N;
  VAR SS REC;
  OUTPUT OUT=B
    MEAN=MEANGRI MEANMSW
    STD=STDGRI STDMSW;
RUN;
PROC SORT DATA=NA; BY N; RUN;
PROC SORT DATA=B; BY N; RUN;
DATA MEAN; MERGE NA B; BY N; RUN;
DATA STREC;
  SET MEAN;
  NORMGRI=(SS-MEANGRI)/STDGRI;
  NORMMSW=(REC-MEANMSW)/STDMSW;
RUN;
PROC PRINT DATA=A; TITLE3 'DATA=A'; RUN;
PROC PRINT DATA=NA; TITLE3 'DATA=NA'; RUN;
PROC PRINT DATA=STREC;
```

```
  TITLE3 'DATA=STREC';
RUN;
****** DATA STEP TO CREATE RANGE OF RECRUITMENTS FOR
       OBSERVED STOCK SIZES.  MUST ENTER STOCK
       SIZES AS DO I=...., RANGE OF RECRUITMENTS
       IS DO J=.... THESE SHOULD INCLUDE ENOUGH
       OF A RANGE TO BRING THE PROBABILITIES
       TO ZERO OR CLOSE TO IT.  THE PREC=J-1 LINE
       IN INCLUDED TO HAVE THE RECRUITMENTS BEGIN
       AT ZERO.;
*******;
DATA V;
****************************
***************************;
  DO I=35.673,46.275,44.545,73.418,64.902,91.580,
       27.743,24.287,50.965,41.588,65.273,80.379,
       25.184,29.707,60.800,117.549,84.816,121.919,
       75.231;
    DO J=1 TO 100;

  *DO I=3,6,9,12;
   * DO J=1 TO 20;
******************************
******************************;
      SS=I;  PREC=J-1;
      OUTPUT V;
    END;
  END;
RUN;
PROC IML;
**********************;
*********************;
  H={.45}; *SMOOTHING PARAMETER MUST BE ENTERED;
*********************;
*********************;
****** NEXT LINES READ IN OBSERVED STOCK
       AND RCRUITMENTS AND A VECTOR OF YEARS;
******;
  USE A VAR {SS REC};
  READ ALL INTO M;
  USE A VAR {YEAR};
  READ ALL INTO YEAR;
PRINT M YEAR;
  MEAN=J(NROW(YEAR),NCOL(M),0);
******** LOOP TO CALCULATE MEAN STOCK SIZE AND
         RECRUITS, LEAVING OUT EACH YEAR IN
         TURN, (CROSS-VALIDATION);
*********;
  DO I=1 TO NROW(YEAR);
     N=(NROW(YEAR))-1;
     IM=J(1,(NROW(YEAR)),1);
     IM[I]=0;
     TIM=T(IM);
     TM=T(M);
     MN=(T(M)*T(IM))/N;
     MEAN[I,]=T(MN);
END;
*PRINT IM TIM TM MN;
PRINT MEAN;
  STD=J(NROW(YEAR),NCOL(M),0);
  V1=J(N,NCOL(MEAN),0);
  V2=J(1,NCOL(MEAN),0);
  V3=J(1,NCOL(MEAN),0);
************* LOOP TO CALCULATE STANDARD
              DEVIATION;
***************;
    DO I=1 TO NROW(YEAR);
     DO J=1 TO N;
      IF I=1 THEN MM=M[(I+1):NROW(YEAR),];
```

```
      IF I > 1 & I < NROW(YEAR) THEN
         MM=M[(1:(I-1)),]//M[((I+1):NROW(YEAR)),];
      IF I=NROW(YEAR) THEN MM=M[(1:(NROW(YEAR)-1)),];
      V1[J,]=((MM[J,]-MEAN[I,])#(MM[J,]-MEAN[I,]));
     END;
*PRINT I J MM V1;
      V2[,]=V1[+,];
      V3[,]=V2/(N-1);
      STD[I,]=V3##.5;
    END;

*PRINT I J V1 V2 V3 STD;
NM1=J(NROW(MM),NCOL(MM),0);
************ LOOP TO NORMALIZE RAW DATA;
*************
*************;
    DO I=1 TO NROW(MEAN);
     DO J=1 TO NROW(MM);
      IF I=1 THEN MM=M[(I+1):NROW(YEAR),];
      IF I > 1 & I < NROW(YEAR) THEN
         MM=M[(1:(I-1)),]//M[((I+1):NROW(YEAR)),];
      IF I=NROW(YEAR) THEN MM=M[(1:(NROW(YEAR)-1)),];
      NM1[J,]=(MM[J,]-MEAN[I,])/(STD[I,]);
     END;
*PRINT MM NM1;
    IF I=1 THEN NORM=NM1;
    IF I > 1 THEN NORM=NORM//NM1;
*PRINT NORM;
    END;
************* READS IN DATA FILE FOR OBSERVED
              STOCK SIZE AND RANGE OF RECRUITS;
*************;
USE V VAR (SS PREC);
READ ALL INTO V;
*PRINT V;
NR=NROW(V)/NROW(YEAR);
NV1=J(NR,NCOL(V),0);
YV1=J(NR,1,0);
*************** LOOP TO NORMALIZE OBSERVED STOCK
                SIZES AND RANGE OF RECRUITS BASED
                ON MEANS AND STDS WHEN A GIVEN
                STOCK SIZE WAS LEFT OUT OF DATA
                SET;
***************;
  DO I=1 TO NROW(MEAN);
       NV=V[(1+(NR#(I-1))):(NR+(NR#(I-1))),];
     DO J=1 TO NR;
       NV1[J,]=(NV[J,]-MEAN[I,])/STD[I,];
     END;
      IF I=1 THEN  NORMV=NV1;
      IF I > 1 THEN NORMV=NORMV//NV1;
*PRINT NV;
  END;
*PRINT NORMV;
E=J(N,NCOL(V),0);
ENV=J(NR,NCOL(V),0);
EN=J(N,NCOL(NORM),0);
F=J(NROW(ENV),1,0);
************** LOOP TO CALCULATE PROBABILITIES ON
               NORMALIZED DATA. ENV= NORMALIZED DATA
               FROM EACH OBSERVED SS AND RANGE OF
               RECRUITS, WITH GIVEN YEAR LEFT OUT
               OF CALCULATIONS.

  DO I=1 TO NROW(MEAN);
       ENV=NORMV[(1+(NR#(I-1))):(NR+(NR#(I-1))),];
       EN=NORM[(1+(N#(I-1))):(N#I),];
*PRINT ENV EN;
     DO J=1 TO NROW(ENV);
```

```
       DO K=1 TO NROW(EN);
          E[K,]=EXP((-.5)#(((ENV[J,]-EN[K,])/H[1,])##2));
       END;
*PRINT E2;
      IF J=1 THEN EX=E;
      IF J > 1 THEN EX=EX//E;
*PRINT E3;
      P=E[,#];
      PS=SUM(P);
      CON=1/((NROW(EN))#((2#3.14159265)##((NCOL(M))/2))#(H[,]));
      Q=CON#PS;
      F[J,]=CON#PS;
*PRINT E P PS CON Q;
     END;
*PRINT F;
     IF I=1 THEN F2=F;
     IF I > 1 THEN F2=F2//F;
  END;
*PRINT F2;
HR=REPEAT(H,NROW(F2),1);
YV1=J(NR,1,0);
 DO I=1 TO NROW(MEAN);
   DO J=1 TO NR;
      YV1[J,]=YEAR[I,];
   END;
  IF I=1 THEN YV=YV1;
  IF I > 1 THEN YV=YV//YV1;
 END;
PRINT YV;
LS=F2||V||YV||HR;
*PRINT LS;
FNAME={'F' 'SS' 'PREC' 'REFYR' 'H'};
CREATE F2V FROM LS[COLNAME=FNAME];
APPEND FROM LS;
QUIT;
PROC PRINT DATA=F2V;
  TITLE3 'DATA=F2V';
RUN;
PROC SORT DATA=F2V; BY REFYR SS; RUN;
PROC SORT DATA=A; BY REFYR SS; RUN;
PROC MEANS DATA=F2V NOPRINT;
  BY REFYR;
  ID SS;
  VAR F;
  OUTPUT OUT=MAX
   MAX=MF;
RUN;
PROC PRINT DATA=MAX; TITLE3 'DATA=MAX'; RUN;
PROC SORT DATA=MAX; BY REFYR SS; RUN;
DATA CALC;
  MERGE F2V MAX A;
  BY REFYR SS;
  IF F=MF THEN DO;
     Y=(PREC-REC)**2;
     VAR+Y;
   END;
   ELSE DELETE;
RUN;
PROC PRINT DATA=CALC;
  VAR H F MF REFYR SS PREC REC Y VAR;
  TITLE3 'DATA=CALC';
RUN;
FILENAME OUTFIL 'DUA1:[CLAYTOR.FORECAST]K1.OUT';
  DATA _NULL_; SET CALC NOBS=COUNT; FILE OUTFIL MOD;
  IF _N_=COUNT THEN PUT H 7.2 VAR 30.5;
RUN;
DATA X;
  INFILE 'DUA1:[CLAYTOR.FORECAST]K1.OUT' MISSOVER;
  INPUT H VARIANCE;
```

```
RUN;
PROC SORT DATA=X OUT=SX;
  BY DESCENDING H;
RUN;
PROC PRINT DATA=SX;
  VAR H VARIANCE;
  FORMAT H 7.2 VARIANCE 30.5;
RUN;
DATA _NULL_; SET SX; FILE 'K.OUT';
  PUT H 7.2 VARIANCE 30.5;
RUN;
```

C. Program for estimating smoothing parameters for each dimension simultaneously using the Gaussian kernel estimator.

```
* MULTIH.sas;
************************************;
title 'MULTIH.SAS MULTIVARIATE H, 1SW-MSW (TOGETHER)';
TITLE2 'MIRAMICHI 1SW YEARS 1971-1989';
  ** SEE T END FOR ADDING TITLE2 LINE TO FINAL OUTPUT;
************************************;
/*
*************************************
   ONE SMOOTHING PARAMETER CALCULATED WITH BOTH VARIABLES
    IN THE MODEL.
   CALCULATES MULTIVARIATE SMOOTHING PARAMETER USING THE MODIFIED
    LIKLIHOOD APPROACH IDENTIFIED IN NOAKES 1989: CJFAS
    46:2046.
   IN DATA STEP RC2 AN EXTRA EXPONENT STEP MUST BE ADDED
    FOR EACH VARIABLE INCLUDED.
   VALUE FOR D, DATA STEP RC2 IS DETERMINED BY ITERATION.
    BEGIN WITH HIGH VALUE OF D THAT APPROACHES INFINITY.
    DECREASE D WITH EACH ITERATION UNTIL VALUE OF D THAT
    MAXIMIZES LIKLIHOOD VARIABLE IS FOUND.
   THE VALUES OF D AND LIKLIHOOD CAN BE SEEN AT THE END
    OF EACH RUN BY TYPING H.OUT.
   AT THE END OF THE JOB A PRINT OF LAST DATA SET DATA=SX
    PROVIDES A RECORD OF ITERATIONS WITH TITLES OF RUN.
   DELETE OLD H.OUT FILES BEFORE BEGINNING A SET OF ITERATIONS
    WITH NEW YEARS OR DATA.  OTHERWISE NEW RESULTS WILL MIX WITH
    OLD.
   CHECK DROP STATEMENT FOR DATA STEP RC TO MAKE SURE YOU ARE
    NOT DROPPING A VARIABLE THAT IS NECESSARY TO KEEP FOR THE
    MODEL YOU ARE EXAMINING.
   CHANGE VALUE OF N FOR EACH RUN IF NECESSARY.
   CHANGE DGRI AND DMSW ETC. TO APPROPRIATE VARIABLE FOR
    SMOOTHING PARAMETERS.
*/

*normal1.sas;
libname a 'dua1:[claytor.forecast]';

data a;
infile 'dua1:[claytor.forecast]miram.dat' missover;
*input for miram.dat;
  input GRYEAR GR SAL SFA2 SFA3 SFA4 RESIDUAL SFA5 SFA6 SFA7
        SFA8 SFA10 SFA12 RES86/ RES87 RES88 RES89;
  GRI=GR/1000;
  MSW=SAL/1000;
  COMB=SUM(SFA5,SFA6,SFA7,SFA8,SFA10);
  YEAR=GRYEAR;
  LGRI=LOG(GRI);
  LMSW=LOG(MSW);
     *TO MAKE DATA CONSISTENT WITH NEWFOUNDLAND CATCH DATA DELETE ALL
         YEARS UP TO 1973;
  *IF YEAR LE 73 THEN DELETE;
       *DELETE YEAR=90 BECAUSE NO RECRUIT ESTIMATES ARE AVAILABLE FOR
           FOR 1991 AS YET;
  IF YEAR=90 THEN DELETE;
    *DELETE MSW YEAR-1 FOR YEAR TRYING TO PREDICT.  IF WANT TO
          FORECAST FOR MSW IN 1989 THEN DELETE YEAR=88;
********************************
********************************;
   *IF YEAR=76 THEN DELETE;
   *IF YEAR GE 86 THEN DELETE;
   N=19;
run;
proc means data=a n mean std MIN MAX;
  ID N;
  var LGRI LMSW  SFA4;
  output out=B
```

```
     n=ngri
     mean=MEANGRI MEANMSW MEANSFA4
     std=STDGRI STDMSW STDSFA4;
run;
/*
proc print data=B;
  title3 'mean and std for input data, data=B';
run;
*/
PROC SORT DATA=A OUT=SA(DROP=RESIDUAL SFA5 SFA6 SFA7 SFA8 SFA10
                             SFA12 RES86 RES87 RES88 RES89);
  BY N;
RUN;
/*
data a;
  infile 'dua1:[claytor.forecast]fake.dat' missover;
  input gryear gri sfa4 msw;
  year=gryear;
  n=4;
run;

proc means data=a n mean std min max;
  id n;
  var gri sfa4 msw;
  output out=b
    n=ngri
    mean=meangri meansfa4 meanmsw
    std=stdgri stdsfa4 stdmsw;
run;
*/
/*
proc print data=a; title3 'data=a'; run;
proc print data=b; title3 'data=b'; run;

proc sort data=a out=sa;
  by n;
run;


PROC SORT DATA=B OUT=SB;
  BY N;
RUN;
DATA MEAN;
  MERGE SA SB;
  BY N;
RUN;
DATA STREC;
  SET MEAN;
*************************
*************************;
   NORMGRI=(LGRI-MEANGRI)/STDGRI;
   NORMMSW=(LMSW-MEANMSW)/STDMSW;
   NORMSFA4=(SFA4-MEANSFA4)/STDSFA4;
   Dgri=NORMgri;
   dmsw=normmsw;
*************************
*************************;

PROC PRINT DATA=MEAN;
  TITLE3 'DATA=MEAN';
RUN;
*/
/*
proc print data=a;
 title3 'Miramichi miram.dat';
  title4 'data=a original data set';
   VAR GR SAL GRI MSW SFA4 LGRI LMSW;
run;
*/
```

```
/*
proc means data=STREC;
  var Dgri dmsw;
run;

DATA DA(DROP=GRI SAL SFA2 SFA4  RESIDUAL SFA5 SFA6 SFA7 SFA8
             SFA9 SFA10 SFA12 RES86 RES87 RES88 RES89
             MEANGRI STDGRI MEANMSW STDMSW NORMGRI NORMMSW
             MEANSFA4 STDSFA4 NORMMSW);
   SET STREC;
RUN;
*/
/*
data da;
  set strec;
run;
*proc print data=da; title3 'data=da'; run;
*/
DATA DA;
  SET A;
  DGRI=LGRI;
  DMSW=LMSW;
RUN;

***************;
****************;
data rc2(DROP=SUMF) totf(drop=recruits f);
  *set rc;
   SET DA;
   REFYR=year;
***************************;
***************************;
   REFgri=Dgri;
   refmsw=dmsw;
   DIM=2;
***************************;
***************************;

**************;
   D=.35;  *KERNEL ESTIMATER;
**************;

cumf=0;
* relate each observation to all the others;
do i=1 to count;
  SET DA POINT=I NOBS=COUNT;
  if i=_n_ then go to nexti;
**********************************
**********************************;
  x=REFgri-Dgri; y=refmsw-dmsw;
  f=(exp(-(x**2)/(2*(D**2))))*(exp(-(y**2)/(2*(d**2))));
**********************************
**********************************;
  cumf+f;
  output rc2;
nexti: end;
con=(1/(n*(d**dim)*((2*3.14159265)**(dim/2))));
sumf=con*cumf;
*sumf=(1/(N*(d**DIM)*((2*3.14159265)**(dim/2))))*cumf; output totf;
run;

*proc print data=rc2; title3 'data set rc2'; run;
proc print data=totf; title3 'data set totf'; run;

proc iml;
  use totf var {sumf D};
  read all into x;
  PRINT X;
  Y=X[,1];
```

```
  Z=X[,1];
  PRINT Y;
  PRINT Z;
  C=1;
 DO I=2 TO NROW(Z);
    Y[I,]=(Y[C,])#(Z[I,]);
    C=C+1;
  END;
PRINT Y;
LIKE=(Y[NROW(Z),])#(10**30);
D=X[1,2];
GK=D||LIKE;
*PRINT LIKE;
*PRINT D;
PRINT GK;
FILENAME OUT 'H.dat';
FILE OUT;
  DO I=1 TO NROW(GK);
     DO J=1 TO NCOL(GK);
        PUT (GK[I,J]) 30.10 +3 @;
     END;
    PUT /;
  END;
CLOSEFILE OUT;
QUIT;
filename outfil 'dua1:[claytor.forecast]h1.out';
data xx;
infile 'dua1:[claytor.forecast]h.dat' missover;
input d like;
 if d=. then delete;
run;
data _null_;
  set xx;
  file outfil mod;
  put d 12.6 +3 like 30.;
run;
data x;
  infile 'dua1:[claytor.forecast]h1.out' missover;
  input d like;
run;
proc sort data=x out=sx;
  by descending d;
run;
proc print data=sx;
  var d like;
  FORMAT D 9.2  LIKE 30.;
  TITLE2 'MIRAM.dat';
  TITLE3 'DATA=SX';
*************************
*************************;
  title4 'Dgri=NORMgri, DMSW=NORMMSW';
  TITLE5 'LN NORMALIZED';
*************************
*************************;
run;
data _null_;
  set sx;
  file 'h.out';
  put d 12.6 +3 like 30.;
run;
```

D. Program to calculate Cauchy step function distribution.

```
* pdf.sas;
******************************;
title 'MIRAMICHI MSW FORECAST 1991';
TITLE2 'CAUCHY DISTRIBUTION - 1SW x MSW';
TITLE3 '1SW YEAR=1990,  1SW YEARS 1971-1990';
TITLE4 'H=17';
******************************;
* Calculates pdf using D estimated from kernal.sas;
*  See Evans and Rice. 1988. Predicting recruitment
    from stock size ... J. Cons. Explor. Mer. 44: 111-122.
    Rice and Evans. 1988.  Tools for embracing uncertainty
    in the management of the cod ..  J. Cons. Explor. Mer. 45:        73-81.
   For details regarding methodology.;
 ************ Must add D value determined from Kernel.sas and
     REFSS=reference stock, spawning stock for
     which you are trying to predict recruits.
     In data rc2 and proc print for data rc3 AND TITLE OF GRAPH;
****************;
options linesize=79;
libname a 'dua1:[claytor.forecast]';
filename pdf 'dua1:[claytor.forecast]mir.gsf';
data a(KEEP=GRYEAR YEAR SS RECRUITS);
infile 'dua1:[claytor.forecast]RESMIR.dat' missover;
 *INPUT FOR RESMIR.DAT;
  INPUT GRYEAR GRI MSW CAUCHY LS LK MEAN GREEN SFA4;
  *IF GRYEAR LE 73 THEN DELETE;
  *IF GRYEAR=89 THEN DELETE;
  SS=GRI/1000;
  RECRUITS=MSW/1000;
  YEAR=GRYEAR;
/*
  *INPUT FOR MIR.DAT;
  *input year ss recgr recsal;
  *recruits=sum(recgr,recsal);

 *INPUT FOR MIR1SW.DAT;
 * INPUT GRYEAR GR SAL SFA2 SFA3 SFA4;
 *input for miram.dat;
  input gryear gr sal sfa2 sfa3 sfa4 residual sfa5 sfa6
        sfa7 sfa8 sfa10 sfa12 RES86/RES87 RES88 RES89;
 *IF GRYEAR LE 73 THEN DELETE;
 *IF GRYEAR GE 86 THEN DELETE;
  SS=GR/1000;
  RECRUITS=SAL/1000;
  YEAR=GRYEAR;
 *DELETE YEAR-1 TO PREDICT, I.E.for  1988 delete ge 87
     if trying to predict a year based on all years
      up to that year MUST USE GE;
************************************;
 * if year ge 83 then delete;
************************************;
  *if trying to predict a year i.e. 1983 based on all years
   except previous year i.e  1982 then delete year-1 FOR 83 DELETE
     82 THE GRILSE YEAR;
**********************************************;
  if year eq 89 then delete;
**********************************************;
 *MUST DELETE YEAR=90 BECAUSE RECRUITS ARE UNKNOWN;
  IF YEAR=90 THEN DELETE;
*/
/*
  *INPUT FOR MIRCOUNT.DAT;
   INPUT GRYEAR E1SW EMSW L1SW LMSW LAB;
    T1SW=SUM(E1SW,L1SW);
    TMSW=SUM(EMSW,LMSW);
    SS=T1SW/100;
```

```
    RECRUITS=TMSW/15;
    YEAR=GRYEAR;
    IF YEAR=80 OR YEAR=84 THEN DELETE;
   * DELETE YEAR TRYING TO PREDICT, I.E. 1989 COUNTS;
    IF YEAR=88 OR YEAR=89 THEN DELETE;
 */

  *INPUT FOR JAKE.DAT;
  * INPUT YEAR SS RECRUITS;

  *input for dempenv.dat
     year=label ss=temperature recruits=smolt counts;
*  input year ss recruits;
run;

proc print data=a;
  title5 'data=a original data set';
run;

proc sort data=a out=rc;
  by recruits;
run;
/*
proc print data=rc;
  title5 'data=rc sorted data set by recruits';
 run;
*/
***************;
****************;
data rc2;
  set rc;

***************;
  D=17; * D, kernel estimater ;
  REFss=90533/1000;  *STOCK SIZE WISH TO PREDICT RECRUITS;
                     *YEAR-1 STOCK SIZE I.E. FOR 1983 USE 1982;
***************;



  x=REFSS-ss;
  f=1/ (1 + (x/d)**2 );
  cumf+f;
run;

proc summary data=rc2;
  var f;
  id d;
  output out=b
     sum=sumf;
run;
proc sort data=b out=sb;
  by d;
run;
proc sort data=rc2 out=src2;
  by d;
run;

data rc3(drop=_type_ _freq_ x d refss); merge src2 sb;
  by d;


FPCT=100*CUMF/SUMF;

run;
/*
 proc print; title5 'data set rc3';
 run;
*/
```

```
DATA X(KEEP=RECRUITS Y);
  SET RC3;
  Y=LAG(FPCT);
  IF Y=. THEN Y=0;
  IF RECRUITS=. THEN DELETE;
RUN;
DATA Y(KEEP=RECRUITS Y);
  SET RC3;
  Y=FPCT;
RUN;
DATA Z;
  SET X Y;
RUN;
PROC SORT DATA=Z OUT=SZ;
  BY Y RECRUITS;
RUN;
/*
*PROC PRINT DATA=X; TITLE5 'DATA=X'; RUN;
*PROC PRINT DATA=Y; TITLE5 'DATA=Y'; RUN;
*PROC PRINT DATA=Z; TITLE5 'DATA=Z'; RUN;
*/
PROC PRINT DATA=SZ;
  TITLE5 'BDATA=SZ PROBABILITIES';
RUN;

*goptions device=tek4010;
%include 'dua1:[claytor]goptlsr.sas';
goptions gsfname=mirpdf;

data prob;
  set rc3;
  symbol1 v='K' F=SPECIAL;
run;

DATA STEP;
  SET SZ;
  LENGTH FUNCTION $8.;
  XSYS='2';
  YSYS='2';
  LINE=1;
  IF Y=0 THEN FUNCTION='MOVE';
    ELSE
      FUNCTION='DRAW';
  X=RECRUITS;
RUN;
PROC PRINT DATA=STEP; TITLE5 'data=step'; RUN;

proc gplot data=prob GOUT=A.mirfore;
   TITLE5 '';
  AXIS1 LABEL=(A=90 'CUMMULATIVE PROBABILITY')
        ORDER=0 TO 100 BY 10
        MINOR=NONE;
  AXIS2 LABEL=('RECRUITS x 1000')
        ORDER=0 TO 60 by 5
        MINOR=NONE;
  PLOT FPCT*RECRUITS/
       VAXIS=AXIS1
       HAXIS=AXIS2
       NAME='FORE91'
       DES='MSW MIRAM FORE 91 1SW,MSW'
        ANNO=STEP;
RUN;


PROC PRINT DATA=RC2; TITLE5 'DATA=RC2'; RUN;
```

E. Program to calculate probability distribution function using Gaussian kernel estimator.

```
*gdf2C.sas;
TITLE 'MIRAMICHI MSW FORECAST 1991';
TITLE2 'LEAST SQUARES H(COMB), RAW NORMALIZED';
TITLE3 '1SW YEAR=90, 1SW YEARS 1971-1990';
*TITLE3 '1SW YEARS 1971-1989';
TITLE4 '1SW - MSW H=0.5';

LIBNAME A 'DUA1:[CLAYTOR.FORECAST]';
/*
   PLOTS THREE DIMENSIONAL PROBABILITY DISTRIBUTIONS, TWO
    DIMENSIONAL PROBABILITIES FOR CONSTANT STOCK SIZE, AND
    CUMMULATIVE PROBABILITY DISTRIBUTIONS FOR CONSTANT
    STOCK SIZE TO PREDICT RECRUITMENT OR OTHER DEPENDENT
    VARIABLES.
   THE D VALUE IS DETERMINED USING E.G MULTIH.SAS AS DESCRIBED IN
    NOAKES 89: CJFAS 46:2046.
   PLACES WHERE VARIABLES OR STEPS MUST BE CHANGED ARE MARKED
    BY A DOUBLE ROW OF ***********.
   ENTER APPROPRIATE VALUES FOR STOCKS AND RECRUITS IN DATA STEP
    A.
   DATA V STEP MUST BE CHANGED TO FIT EXPECTED STOCK AND
    RECRUITMENTS SO THAT PROBABILITIES WILL GO TO ZERO
    AT EXTREMES.
   VECTOR H MUST BE CHANGED TO THAT APPROPRIATE SMOOTHING
    PARAMETERS FROM GKA.SAS ARE INPUT.  Hs MUST BE ENTERED
    IN SAME ORDER AS THEIR VARIABLES IN THE DATA SET.
    IN THIS EXAMPLE STOCK SIZE(GRI) IS IN THE FIRST COLUMN
    SO ITS H COMES FIRST.  RECRUITS(MSW) ARE IN THE SECOND
    COLUMN SO ITS H COMES SECOND.
   STOCK SIZE IN DATA YPDF MUST BE CHANGED TO THE STOCK
    SIZE THAT IS BEING HELD CONSTANT.
   A SET OF COMMANDS FOR PLOTTING MULTIPLE LINES ON ONE
    GRAPH IS ALSO AVAILABLE.
*/
*FROM NORMAL1.SAS;
data a;
infile 'dua1:[claytor.forecast]miram.dat' missover;
*input for miram.dat;
  input GRYEAR GR SAL SFA2 SFA3 SFA4 RESIDUAL SFA5 SFA6 SFA7
        SFA8 SFA10 SFA12 RES86/ RES87 RES88 RES89;
  GRI=GR/1000;
  MSW=SAL/1000;
  COMM=SFA4;
  YEAR=GRYEAR;
     *TO MAKE DATA CONSISTENT WITH NEWFOUNDLAND CATCH DATA DELETE ALL
         YEARS UP TO 1973;
  *IF YEAR LE 73 THEN DELETE;
       *DELETE YEAR=90 BECAUSE NO RECRUIT ESTIMATES ARE AVAILABLE FOR
           FOR 1991 AS YET;
  IF YEAR=90 THEN DELETE;
    *DELETE MSW YEAR-1 FOR YEAR TRYING TO PREDICT.  IF WANT TO
          FORECAST FOR MSW IN 1989 THEN DELETE YEAR=88;
 *****************************
*****************************;
*  IF YEAR=89 THEN DELETE;      *to delete single years and keep rest;
  * IF YEAR GE 86 THEN DELETE;  *to look at all years up to a year;
   N=19;
*****************************
*****************************;
run;
proc means data=a n mean std STDERR MIN MAX;
  ID N;
  var GR SAL GRI MSW;
  output out=B
     n=ngr
     mean=meangr meansal MEANGRI MEANMSW
```

```
    std=stdgr stdsal STDGRI STDMSW;
run;
proc print data=a;
  title5 'original data=a';
  var gryear gri msw sfa4;
run;
/*
proc print data=B;
  title5 'mean and std for input data, data=B';
run;
*/
PROC SORT DATA=A OUT=SA(DROP=RESIDUAL SFA5 SFA6 SFA7 SFA8 SFA10
                              SFA12 RES86 RES87 RES88 RES89);
  BY N;
RUN;
PROC SORT DATA=B OUT=SB;
  BY N;
RUN;
DATA MEAN;
  MERGE SA SB;
  BY N;
RUN;
DATA STREC;
  SET MEAN;
***********************************
***********************************;
  NORMGRI=(GRI-MEANGRI)/STDGRI;
  NORMMSW=(MSW-MEANMSW)/STDMSW;
  SS=NORMGRI;
  RECRUITS=NORMMSW;
  MEANSS=MEANGRI;
  STDSS=STDGRI;
  MEANREC=MEANMSW;
  STDREC=STDMSW;
***********************************
***********************************;
RUN;
/*
PROC PRINT DATA=MEAN;
  TITLE5 'DATA=MEAN';
RUN;

proc print data=strec;
  title5 'data=strec';
run;
*/
PROC MEANS DATA=STREC;
  VAR NORMGRI NORMMSW;
RUN;
**********************
**********************;
********DATA STEP TO PRODUCE ALL POSSIBLE STOCK AND RECRUITS;
**************************************
**************************************;
DATA V;
  *DO I=1 TO 200 BY 5;
     *STOCK SIZE, LOOP NEEDED TO PRODUCE ALL POSSIBLE COMBINATIONS;
   DO I=90.533;
     *STOCK SIZE FOR 1SW YEAR FORECASTING MSW YEAR+1;
        DO J=1 TO 80 BY 1;        *RECRUITS;
         R=I; C=J-1;       *FOR IND YEARS;
       * R=I-1;  *C=J-1;
   *USE -1 FORMAT IF STARTING VALUES OF 0 ARE NEEDED;
        *R=I; *C=J;
   *USE THIS FORMAT IF START AT VALUE OTHER THAN ZERO;
        PUT R 5.0 C 5.0;  *R=STOCK SIZE, C=RECRUITS;
        OUTPUT V;
     END;
   END;
```

```
RUN;
************************
************************;
/*
PROC PRINT DATA=V;
  TITLE5 'DATA=V';
RUN;
*/
proc iml;
  use STREC var {ss recruits};   *NORMALIZED STOCK AND RECRUITS;
  read all into m;
  d=ncol(m);
  n=nrow(m);
*************************
*************************;
  h={.5 .5};                     *H VALUES DETERMINED USING GKA.SAS;
                                   * H1=STOCK, H2=RECRUITS;
*************************
*************************;
**** CONSTANT IN FORMULA (5) NOAKES 89;
  con=1/((nrow(m))#((2#3.14159265)##(d/2))#h[1]#h[2]);

  use STREC var {meanss meanrec};
  read all into nmx;
  use STREC var {stdss stdrec};
  read all into nsx;
  nm=nmx[1,];            *MEAN VALUES;
  ns=nsx[1,];            *STD VALUES;
*print M nm ns;
USE V VAR {R C};
READ ALL INTO V;         *RAW DATA INPUT FOR ALL POSSIBLE STOCK-REC;
VN=J(NROW(V), NCOL(V), 0);
***LOOP TO CALCULATE NORMALIZED VALUES FOR ALL POSSIBLE STOCK-REC;
DO I=1 TO NROW(V);
   DO J=1 TO NCOL(V);
     VN[I,J]=(V[I,J]-NM[1,J])/NS[1,J];
   END;
END;
*print V vn;

  e=j(nrow(m),ncol(m),0);
  F=J(NROW(VN),1,0);
****LOOP TO CALCULATE PRODUCT PORTION OF FORMULA (5) NOAKES 89;
DO K=1 TO NROW(V);    ***PICKS OUT INDIVIDUAL ROWS OF NORMALIZED MATRIX DATA;
  VV=VN[K,];
*PRINT VV;
   do i=1 to nrow(m);
      do j=1 to ncol(m);
         e[i,j]=exp((-.5)#(((VV[,j]-m[i,j])/h[j])##2));  *PRODUCT PORTION (5);
      end;
   end;


p=(e[,1])#(e[,2]);   *PRODUCT OF E[I,J];
ps=sum(p);           *SUM OF PRODUCT OF E[I,J];
F[K]=CON#PS;
*PRINT P PS F;
END;

PLOTNORM=VN||F;      *VN, NORMALIZED VALUES AND F, PROBABILITIES;
PLOT=V||F;           *V, RAW DATA VALUES AND F, PROBABILITIES FOR PLOTS;
*PRINT PLOTNORM PLOT;
*print CON  F;
*print m n d v h  con;
*print nmx nsx nm ns vn;
****MAKES DATA FILE FOR PLOTTING THREE DIMENSIONAL GRAPHS FROM PLOT MATRIX;
FILENAME OUT 'PLOT.DAT';
FILE OUT;
  DO I=1 TO NROW(PLOT);
```

```
    DO J=1 TO NCOL(PLOT);
      PUT (PLOT[I,J])  9.7 +2 @;
     END;
    PUT /;
  END;
CLOSEFILE OUT;
quit;


DATA PL;
  INFILE 'DUA1:[CLAYTOR.FORECAST]PLOT.DAT' MISSOVER;
  INPUT SS RECRUITS F;
  IF SS=. THEN DELETE;
  *IF SS=0 THEN F=0;     *IF STOCK SIZE IS ZERO THEN PROB OF RECRUITS IS 0;
RUN;

proc print data=pl;
  title5 'data=pl';
run;


FILENAME GDF 'DUA1:[CLAYTOR.FORECAST]GDF.GSF';
*GOPTIONS DEVICE=TEK4010;
%INCLUDE 'DUA1:[CLAYTOR]GOPTLSR.SAS';
GOPTIONS GSFNAME=GDF;
************************
************************;
/*

**** THREE DIMENSIONAL PLOT;
PROC G3D DATA=PL GOUT=A.MIRFORE;
  TITLE5 ' ';
  PLOT RECRUITS*SS=F/
      TILT=45
      ROTATE=-135
      NAME='LSPDF'
      DES='LEAST SQUARE PDF 1SW MSW ROTATE-135';
RUN;
PROC G3D DATA=PL GOUT=A.MIRFORE;
  TITLE5 ' ';
  PLOT RECRUITS*SS=F/TILT=45
  name='LSPDF'
  DES='LEAST SQUARE PDF 1SW MSW NO ROTATE';
RUN;
*/
SYMBOL V=NONE I=SPLINES L=1;
PROC GPLOT DATA=PL GOUT=A.MIRFORE;
  TITLE5 '';
  PLOT F*RECRUITS/
    NAME='LS91'
    DES='LEAST SQUARE RAW NORM FORE 91';
RUN;

****  PICKS OFF STOCK SIZE TO BE HELD CONSTANT FOR
      PDF AND CDF PLOTS TO FORECAST RECRUITS;
data ypdf;
   set pl;
*************************
*************************;

/*
   if ss=35 OR SS=60 OR SS=85 OR SS=110;
   IF SS=35 THEN SS=37.5;
   IF SS=60 THEN SS=62.5;
   IF SS=85 THEN SS=87.5;
   IF SS=110 THEN SS=112.5;
**************************
**************************;
 * MULTIPLE LINES;
```

```
   IF SS=37.5 THEN DO; FA=F; RA=RECRUITS; END;
   IF SS=62.5 THEN DO; FB=F; RB=RECRUITS; END;
   IF SS=87.5 THEN DO; FC=F; RC=RECRUITS; END;
   IF SS=112.5 THEN DO; FD=F; RD=RECRUITS; END;

************************
************************;
*/
run;
/*
PROC PRINT DATA=YPDF;
  TITLE5 'DATA=YPDF';
RUN;
*/
/*
SYMBOL1 I=SPLINES L=1;
SYMBOL2 I=SPLINES L=2;
SYMBOL3 I=SPLINES L=3;
SYMBOL4 I=SPLINES L=7;
** PDF PLOT;

PROC GPLOT DATA=yPdf GOUT=A.MIRFORE;
  TITLE5 ' ';
  PLOT F*RECRUITS;
************************
************************;

 *MULTIPLE LINES;
   PLOT FA*RECRUITS=1
        FB*RECRUITS=2
        FC*RECRUITS=3
        FD*RECRUITS=4/OVERLAY
        NAME='FLEVEL'
        DES='FOUR LEVEL NOAKES IND PROB';
************************
************************;

RUN;
*/

****  CALCULATES CUMMULATIVE PROBALITY PERCENTAGES FOR CUMMULATIVE
      STOCK SIZE AGAINST RECRUITS;

PROC SORT DATA=YPDF OUT=SYPDF;
  BY SS;
RUN;

data acdf;
  set Sypdf;
  BY SS;
  IF FIRST.SS THEN CUMF=0;
  cumf+f;
run;
proc summary data=acdf NWAY;
  CLASS SS;
  var f;
  output out=bcdf
   sum=sumf;
run;
proc sort data=bcdf out=sbcdf;
  by ss;
run;
proc sort data=acdf out=sacdf;
  by ss;
run;
data ycdf(drop=_type_ _freq_);
   merge sbcdf sacdf;
   by ss;
   fpct=100*cumf/sumf;
```

```
/*
********************************
********************************;
 *LINES ADDED FOR MULTIPLE LINES ON SAME GRAPH;
   IF SS=37.5 THEN DO; FA=FPCT; RA=RECRUITS; END;
   IF SS=62.5 THEN DO; FB=FPCT; RB=RECRUITS; END;
   IF SS=87.5 THEN DO; FC=FPCT; RC=RECRUITS; END;
   IF SS=112.5 THEN DO; FD=FPCT; RD=RECRUITS; END;
*************************
*************************;
*/

run;
********************************
********************************;
run;
proc print data=ycdf;
  title5 'data=ycdf';
run;
*** CDF PLOT;

/*
proc gplot data=ycdf GOUT=A.MIRFORE;
   TITLE5 ' ';
   plot fpct*recruits;   *SINGLE GRAPH LINE;
**************************
**************************;

  *MULTIPLE LINES;
  PLOT FA*RA=1
       FB*RB=2
       FC*RC=3
       FD*RD=4/OVERLAY

       NAME='CUMLEVEL'
       DES='FOUR LEVELS NOAKES';

*************************
*************************;
run;
*/
```