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

Canadian Atlantic Fisheries
Scientific Advisory Committee

CAFSAC Research Document 91/22

Ne pas citer sans
autorisation des auteurs[1]

Comité scientifique consultatif
des pêches canadiennes dans
l'Atlantique

CSCPCA Document de recherche
91/22

Forecasts of MSW salmon returns to the Saint John River using non-parametric and parametric models

by

C.J. Harvie and P.G. Amiro

Biological Sciences Branch
Department of Fisheries and Oceans
P.O. Box 550
Halifax, N.S.
B3J 2S7

[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'études 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

Parametric regression and probability density non-parametric regression analyses of multi-sea-winter Atlantic salmon forecasts for Saint John River are compared for precision and accuracy. Comparisions were made between models using only 1SW returns (year i-1) as the predictor and models using 1SW and mean fork length of 1SW fish as the predictors. Non-parametric models involving two independent variables were made in two steps; i.e. the residuals of MSW returns on 1SW returns were predicted using mean fork length. The inclusion of mean fork lengths greatly improved the accuracy of non-parametric based forecasts which used untransformed data. Precision, the 90% prediction interval, was better for parametric modelling but did not always include the observed value (10% of observations). A combined indicator of both precision and accuracy indicated that the three-variable parametric model using log-transformed MSW returns performed best. Further examination of the robustness of the parameter estimates to the data is suggested before committing to one modelling technique.

Résumé

Des analyses de régression paramétriques et des analyses de régression non paramétriques des densités de probabilité des prévisions de retours de saumons de l'Atlantique après plusieurs hivers en mer (MSW) dans la rivière Saint-Jean sont comparées quant à la précision et à l'exactitude. Des modèles n'utilisant que les retours après un seul hiver (1SW) (année i = 1) comme variable prédictive ont été comparés à des modèles utilisant les retours 1SW et la longueur moyenne à la fourche des poissons comme variables prédictives. Les modèles non paramétriques faisant intervenir deux variables indépendantes ont été appliqués en deux étapes; c'est-à-dire que les résidus des retours MSW sur les retours 1SW ont été prévus au moyen de la longueur moyenne à la fourche. L'inclusion des longueurs moyennes à la fourche a de beaucoup amélioré l'exactitude des prévisions non paramétriques faites à partir de données non transformées. La précision, pour un niveau de confiance de 90 %, était meilleure dans le cas de la modélisation paramétrique, la valeur observée n'étant pas toujours incluse (10 % des observations). Un indicateur combiné de la précision et de l'exactitude indiquait que le modèle paramétrique à trois variables faisant intervenir le logarithme des retours MSW était celui qui fonctionnait le mieux. Il est suggéré d'effectuer un examen plus poussé de la robustesse des estimations de paramètres par rapport aux données avant de retenir une méthode particulière de modélisation.

Introduction

The ability to accurately forecast Atlantic salmon (*Salmo salar*) returns to a river is a desirable goal of Atlantic salmon management. Accurate forecasts allow the setting of pre-season harvest levels allowing greater flexibility in allocations while providing a measure of certainty in attaining target spawning escapements. Parametric forecast models of multi sea-winter (MSW) salmon returns, the major contributor to egg depositions, have been based on returns of one sea-winter (1SW) salmon the year previous.

Parametric modelling relies on assumptions concerning error distribution and response pattern similar to a theoretical distribution. Non-parametric frequency distributions are unrestricted by a theoretical distribution and can be fitted to sample data with any form of error distribution. Non-parametric models have been suggested as better suited to fishery models, for which the response pattern is generally unknown, by Rice and Evans (1988) and Evans and Rice (1988) and utilized by Noakes (1989) for *Oncorhynchus nerka*. The performance of non-parametric techniques in current fishery forecasts is of interest to fishery scientists seeking improved forecasting techniques that remain generally understandable.

Non-parametric probability functions may have intuitive appeal in simplifying probability statements concerning the outcome of events but they also require a thorough understanding of their fitting and use. Because the outcome of these analyses must be understood by practitioners, managers and interested clientele, the adoption of a new analytical method requires that significant benefits be attributed to the change.

This paper compares the application of non-parametric probability functions with parametric forecasting models of MSW returns to the Saint John River, New Brunswick. Forecast model variables are those of Ritter et al. (1990) and include the mean length of 1SW fish and the number of 1SW returns the previous year. Comparisons are made of the accuracy and robustness of the forecasts, width of the confidence limits, and residual error. A new method of model evaluation is presented which weights precision and model performance equally. Results are derived for individual years.

Methods

Two types of predictive models, the non-parametric probability density function model described by Noakes (1989) and the parametric regression model, were compared to determine their effectiveness in forecasting accurate returns of MSW salmon in year i (i = 1971 to 1990) to the Saint John River, N.B., using 1SW salmon returns in year i-1 and/or mean fork length of 1SW salmon in year i-1 (Table 1). Transformations of MSW returns in the non-parametric models were made only for the purpose of comparison with the corresponding parametric models.

Model 1 involved constructing the joint probability density function of MSW salmon returns and 1SW salmon returns from the previous year. Model 2 used the natural logarithm of the MSW salmon returns.

Models 3 and 4 were analogous to models 1 and 2 but included the independent variable mean fork length of 1SW salmon returns in a two-step joint probability density function. Silverman (1986) indicated that a sample size of 67 observations is required to ensure that the relative mean square error at a point is at least 0.1 for estimating a density model with three dimensions (variables). Therefore, since the models in this paper use only 20 observations each, the joint probability density function was constructed in two steps, each using only two variables. The first step consisted of constructing the joint probability density function of MSW salmon returns and 1SW salmon returns (ie. models 1 and 2). The second step involved constructing the joint probability density function of the residuals from step one and the mean fork length of 1SW grilse returns. The forecasted residuals from the second step were then added to the forecasted returns from the first step to produce new return forecasts.

Models A and B were parametric regression models using the same variables as models 3 and 4, respectively.

In all models, the data for all years were used to determine the model parameters. Forecasts of MSW salmon returns were made for each year.

The procedure used to estimate the set of multivariate smoothing parameters for the probability density function model follows that given by Noakes (1989). Data were normalized to deal with variables of different scales and weighted to decrease the influence of outliers on the likelihood function. Weighting for a model with two variables (i.e., MSW salmon returns and 1SW salmon returns) used the square of the distance between the X and Y coordinates of the normalized data point (Figure 1). Normalized points with a squared distance of 1 or less (i.e., within one standard deviation) contribute a weight of 1 to the likelihood function, whereas normalized points with a squared distance greater than 1 contribute a weight equal to the inverse of the squared distance.

Each model was assessed for accuracy using several techniques: the width of the 90% prediction interval (PI) (Snedecor and Cochran, 1989), the residual sum of squares, and a combination of these two statistics (percent PI + percent SS) as

$$(\%PI + \%SS)_{ij} = \left(\frac{PI\ width}{\sum\limits_{i,j} (PI\ width)_{ij}} + \frac{Residual\ SS}{\sum\limits_{i,j} (Residual\ SS)_{ij}} \right) x\ 100$$

where i = 1971,...,1990 and
j = model

Percent PI + Percent SS is a new method of model evaluation designed to give equal weight to both the prediction interval width and the residual sum of squares by expressing each component as a percentage of its total over all models and years.

Results and Discussion

The set of multivariate smoothing parameters was estimated for each of the probability density models (Table 2). The parametric regression models produced significant regression coefficients ($P < 0.001$) for the unlogged and the semi-logged models, with adjusted R^2 values of 0.55 and 0.60, respectively.

Non-parametric Models 1 and 2 - two variables in one step

Model 1, involving only MSW salmon returns and 1SW salmon returns, produced the smoothed probability density function shown in Figure 2. Forecasts of MSW returns varied little over the years 1971 to 1990 (Table 3a). This analysis resulted in large values for residual sum of squares and percent PI + percent SS (Table 4). The wide scatter of the data points resulted in wide prediction intervals on all forecasts.

Using the natural logarithms of MSW salmon returns (model 2) produced the smoothed probability density function shown in Figure 3. This model slightly decreased the differences between the actual and forecasted MSW returns compared to the unlogged model, but it increased the prediction interval widths (Tables 3a and 4). Use of the mean MSW returns as forecasts produced a residual sum of squares similar to those of models 1 and 2.

Non-parametric Models 3 and 4 - three variables in two steps

Figure 4 shows a plot of MSW salmon returns versus 1SW salmon returns, where the size of each point is proportional to the mean fork length of 1SW salmon returns. The smoothed probability density functions for each of the two steps, involving MSW salmon returns, 1SW salmon returns and mean fork length, are shown in Figures 2 and 5. Inclusion of the independent variable, mean fork length, in model 3 produced forecasted MSW salmon returns closer to the actual returns and gave a much smaller value for the residual sum of squares than did the mean of MSW salmon returns or models 1 and 2 (Tables 3b and 4).

Figures 3 and 6 show the smoothed probability density functions calculated from the natural logarithm of MSW salmon returns, 1SW salmon returns and mean fork length for each of the two steps. Taking the natural logarithm of MSW salmon returns increased the difference between the actual and forecasted returns as compared to the unlogged model. As well, the residual sum of squares increased, compared to the two-variable model (model 1), indicating that the mean fork length of 1SW salmon returns did not explain or improve the initial MSW forecast.

Both models produced larger values for prediction interval width and percent PI + percent SS because of the additive effect of prediction intervals in the two-step method. Therefore, comparisons of models 3 and 4 to other models concentrated on residual sum of squares values as an indication of forecasting accuracy. In general, model 3 produced more accurate forecasts of MSW salmon returns, by including the variable for mean fork length, than did the mean of MSW salmon returns or models 1, 2 or 4.

Parametric Models A and B

Models A and B, the parametric regression models, gave smaller values for the residual sum of squares, percent PI + percent SS, and prediction interval width (Tables 3c and 4). Taking the natural logarithm of MSW salmon returns produced more accurate forecasts of returns than the unlogged model. However, the 1975 forecast in the logged model and the 1983 forecasts in both the unlogged and logged models produced prediction intervals which did not include the actual MSW salmon return (Tables 3c and 4). This result is within that expected for 90% prediction intervals with 20 observations.

In general, model B performed better than all other models in all techniques, with model A giving similar results.

Robustness of the parameter estimates to the data may provide further insight into the selection of the two types of models examined. Estimates of fully jacknifed parameters and predictions made without the data point included in the parameter estimates may be worthy of examination.

References

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.

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

Rice, J.C., and 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.

Ritter, J.A., T.L. Marshall and P.R. Boudreau. MS 1990. Model development for predicting multi-sea-winter salmon returns to the Saint John River, New Brunswick. CAFSAC Res. Doc. 90/84:280.

Silverman, B.W. 1986. Density estimation for statistics and data analysis. (Monographs on Statistics and Applied Probability). Chapman and Hall, London, 175 p.

Snedecor, G.W., and W.G. Cochran. 1989. Statistical methods. The Iowa State University Press, Eighth Edition, Ames, 503 p.

Table 1. MSW salmon returns, 1SW salmon returns, and mean fork length of 1SW salmon returns to the Saint John River, 1971 to 1990.

Year i	MSW salmon in year i	1SW salmon in year i-1	Mean 1SW fork length (cm)
1971	4733	3057	54.6
1972	4899	1709	55.7
1973	2518	908	57.0
1974	5811	2070	54.6
1975	7441	3656	56.1
1976	8177	6858	55.5
1977	9712	8147	55.4
1978	4021	3977	56.0
1979	2754	1902	56.3
1980	10924	6828	55.6
1981	5991	8482	58.1
1982	5001	5782	56.3
1983	3447	4958	55.3
1984	9779	4309	55.3
1985	10436	8311	55.6
1986	6128	6526	55.7
1987	4352	7904	57.6
1988	2652	5909	58.0
1989	4072	8930	58.6
1990	3329	9522	59.1

Table 2. Models for forecasting MSW salmon returns to the Saint John River, 1971 to 1990.

Model	Model Type	Variables	Parameters
1	Probability density function	MSW returns 1SW returns	H1 = 0.62 H2 = 0.72
2	Probability density function	Ln(MSW returns) 1SW returns	H1 = 0.72 H2 = 0.64
3	Probability density function	Model 1 residuals Mean fork length	H1 = 0.73 H2 = 0.41
4	Probability density function	Model 2 residuals Mean fork length	H1 = 0.28 H2 = 0.53
A	Parametric regression	MSW returns 1SW returns Mean fork length	Constant = 97782.209 Coeff. = 0.720 Coeff. = -1703.151
B	Parametric regression	Ln(MSW returns) 1SW returns Mean fork length	Constant = 25.019 Coeff. = 0.128E-3 Coeff. = -0.305

Table 3a. MSW salmon forecast returns, residuals (actual - forecast), standardized residuals, prediction interval (PI) widths, and percent prediction interval (%PI) width + percent residual sum of squares (%SS) (see text) for the mean and models 1 and 2 of MSW salmon returns to the Saint John River, 1971 to 1990.

		Mean=5809		Model 1					Model 2				
MSW year	MSW return	Residual	Std. residual	Forecast return	Residual	Std. residual	PI width	%PI+ %SS	Forecast return	Residual	Std. residual	PI width	%PI+ %SS
1971	4733	-1076	-0.3966	4434	299	-0.3923	9253	0.68	4786	-53	-0.2145	9391	0.67
1972	4899	-910	-0.3354	4275	624	-0.2693	8514	0.67	4699	200	-0.1134	8126	0.59
1973	2518	-3291	-1.2132	4117	-1599	-1.1111	8038	1.01	4531	-2013	-0.9977	7423	1.22
1974	5811	2	0.0008	4328	1483	0.0560	8672	0.99	4742	1069	0.2339	8444	0.80
1975	7441	1632	0.6017	4434	3007	0.6331	9466	2.22	4742	2699	0.8853	9977	1.96
1976	8177	2368	0.8730	4646	3531	0.8316	10206	2.86	6824	1353	0.3474	11907	1.16
1977	9712	3903	1.4389	4593	5119	1.4329	10153	5.20	5809	3903	1.3664	12060	3.46
1978	4021	-1788	-0.6591	4434	-413	-0.6620	9624	0.72	4742	-721	-0.4814	10273	0.82
1979	2754	-3055	-1.1262	4328	-1574	-1.1016	8619	1.04	4742	-1988	-0.9878	8275	1.26
1980	10924	5115	1.8857	4646	6278	1.8718	10206	7.45	6763	4161	1.4695	11907	3.80
1981	5991	182	0.0671	4540	1451	0.0439	10153	1.08	5147	844	0.1440	11905	0.97
1982	5001	-808	-0.2978	4540	461	-0.3310	10152	0.76	5486	-485	-0.3871	11523	0.86
1983	3447	-2362	-0.8707	4434	-987	-0.8793	9941	0.88	4918	-1471	-0.7812	10932	1.15
1984	9779	3970	1.4636	4434	5345	1.5185	9730	5.57	4786	4993	1.8020	10482	5.00
1985	10436	4627	1.7058	4593	5843	1.7071	10153	6.55	5387	5049	1.8243	12037	5.21
1986	6128	319	0.1177	4646	1482	0.0556	10206	1.10	6403	-275	-0.3032	11821	0.86
1987	4352	-1457	-0.5371	4646	-294	-0.6169	10206	0.74	7207	-2855	-1.3342	12060	2.25
1988	2652	-3157	-1.1638	4540	-1888	-1.2205	10152	1.33	5637	-2985	-1.3862	11628	2.35
1989	4072	-1737	-0.6403	4487	-415	-0.6627	10153	0.75	4742	-670	-0.4611	11928	0.93
1990	3329	-2480	-0.9142	4381	-1052	-0.9040	10153	0.91	4409	-1080	-0.6249	11691	1.03

Table 3b. MSW salmon forecast returns, residuals (actual - forecast), standardized residuals, prediction interval (PI) widths, and percent prediction interval (%PI) width + percent residual sum of squares (%SS) (see text) for models 3 and 4 of MSW salmon returns to the Saint John River, 1971 to 1990.

		Model 3					Model 4				
MSW year	MSW return	Forecast return	Residual	Std. residual	PI width	%PI+ %SS	Forecast return	Residual	Std. residual	PI width	%PI+ %SS
1971	4733	5014	-281	-0.0965	19195	1.38	5075	-342	-0.7427	16416	1.19
1972	4899	5321	-422	-0.1772	18922	1.38	4305	594	-0.3168	15590	1.17
1973	2518	3247	-729	-0.3531	16272	1.25	2332	186	-0.5024	14838	1.06
1974	5811	4908	903	0.5817	18614	1.47	5031	780	-0.2321	15469	1.21
1975	7441	4911	2530	1.5137	19874	2.51	4250	3191	0.8649	17586	2.99
1976	8177	9317	-1140	-0.5885	20563	1.69	6527	1650	0.1637	19322	1.84
1977	9712	9316	396	0.2913	20509	1.49	5561	4151	1.3017	19426	4.33
1978	4021	5066	-1045	-0.5341	20032	1.62	4299	-278	-0.7135	17834	1.29
1979	2754	4494	-1740	-0.9322	18923	1.87	4201	-1447	-1.2454	15982	1.50
1980	10924	9162	1762	1.0738	20615	2.00	6417	4507	1.4636	19322	4.85
1981	5991	4085	1906	1.1563	17661	1.88	2314	3677	1.0860	17027	3.52
1982	5001	4706	295	0.2335	20456	1.47	4945	56	-0.5616	19230	1.37
1983	3447	9209	-5762	-3.2360	20297	7.11	4719	-1272	-1.1658	18250	1.58
1984	9779	9209	570	0.3910	20086	1.49	4587	5192	1.7753	17800	5.87
1985	10436	9109	1327	0.8246	20562	1.77	5041	5395	1.8677	19452	6.36
1986	6128	5692	436	0.3142	20614	1.50	6009	119	-0.5329	19285	1.38
1987	4352	3983	369	0.2759	17611	1.28	4520	-168	-0.6635	17377	1.24
1988	2652	4085	-1433	-0.7563	17608	1.61	2804	-152	-0.6562	16799	1.20
1989	4072	3980	92	0.1172	17351	1.24	3909	163	-0.5129	16806	1.20
1990	3329	3615	-286	-0.0993	16781	1.21	3527	-198	-0.6771	15984	1.15

Table 3c. MSW salmon forecast returns, residuals (actual - forecast), standardized residuals, prediction interval (PI) widths, and percent prediction interval (%PI) width + percent residual sum of squares (%SS) (see text) for models A and B of MSW salmon returns to the Saint John River, 1971 to 1990.

		Model A					Model B				
MSW year	MSW return	Forecast return	Residual	Std. residual	PI width	%PI+ %SS	Forecast return	Residual	Std. residual	PI width	%PI+ %SS
1971	4733	6991	-2258	-1.3142	6792	1.35	6357	-1624	-1.1711	7356	0.97
1972	4899	4147	752	0.4398	6791	0.58	3825	1074	0.4476	4425	0.51
1973	2518	1356	1162	0.6787	7286	0.75	2322	196	-0.0791	2904	0.21
1974	5811	6281	-470	-0.2723	6849	0.53	5602	209	-0.0713	6542	0.47
1975	7441	4868	2573	1.5009	6547	1.60	4344	3097	1.6614	4829*	1.98
1976	8177	8195	-18	-0.0089	6713	0.48	7858	319	-0.0053	8977	0.66
1977	9712	9293	419	0.2457	6977	0.53	9555	157	-0.1025	11388	0.82
1978	4021	5269	-1248	-0.7257	6521	0.73	4666	-645	-0.5837	5163	0.44
1979	2754	3264	-510	-0.2956	6830	0.53	3265	-511	-0.5033	3801	0.32
1980	10924	8003	2921	1.7037	6679	1.93	7593	3331	1.8018	8625	2.51
1981	5991	4936	1055	0.6163	6789	0.67	4377	1614	0.7716	5063	0.81
1982	5001	6058	-1057	-0.6144	6475	0.65	5365	-364	-0.4151	5892	0.44
1983	3447	7168	-3721	-2.1668	6602*	2.83	6549	-3102	-2.0578	7346*	2.17
1984	9779	6700	3079	1.7958	6589	2.09	6027	3752	2.0544	6746	2.88
1985	10436	9071	1365	0.7970	6933	0.81	9180	1256	0.5568	10865	1.04
1986	6128	7615	-1487	-0.8650	6615	0.85	7086	-958	-0.7715	7966	0.72
1987	4352	5372	-1020	-0.5928	6648	0.65	4735	-383	-0.4265	5352	0.41
1988	2652	3254	-602	0.3493	6765	0.54	3247	-595	-0.5537	3741	0.33
1989	4072	4407	-335	-0.1937	6965	0.52	3980	92	-0.1415	4735	0.34
1990	3329	3982	-653	-0.3790	7191	0.59	3686	-357	-0.4109	4543	0.35

* Prediction interval does not include actual MSW salmon return.

Table 4. Average prediction interval widths, residual sums of squares, and average percent PI + percent SS (see text) for the mean and each model for MSW salmon returns to the Saint John River, 1971 to 1990.

Model	Prediction interval width (PI)	Residual sum of squares (SS) (x 10^6)	Percent PI + percent SS
Mean		139.808	
1	9688	168.138	2.13
2	10690	123.658	1.82
3	19127	58.162	1.86
4	17490	125.068	2.32
A	6778	55.953	0.96
B	6313	54.933	0.92

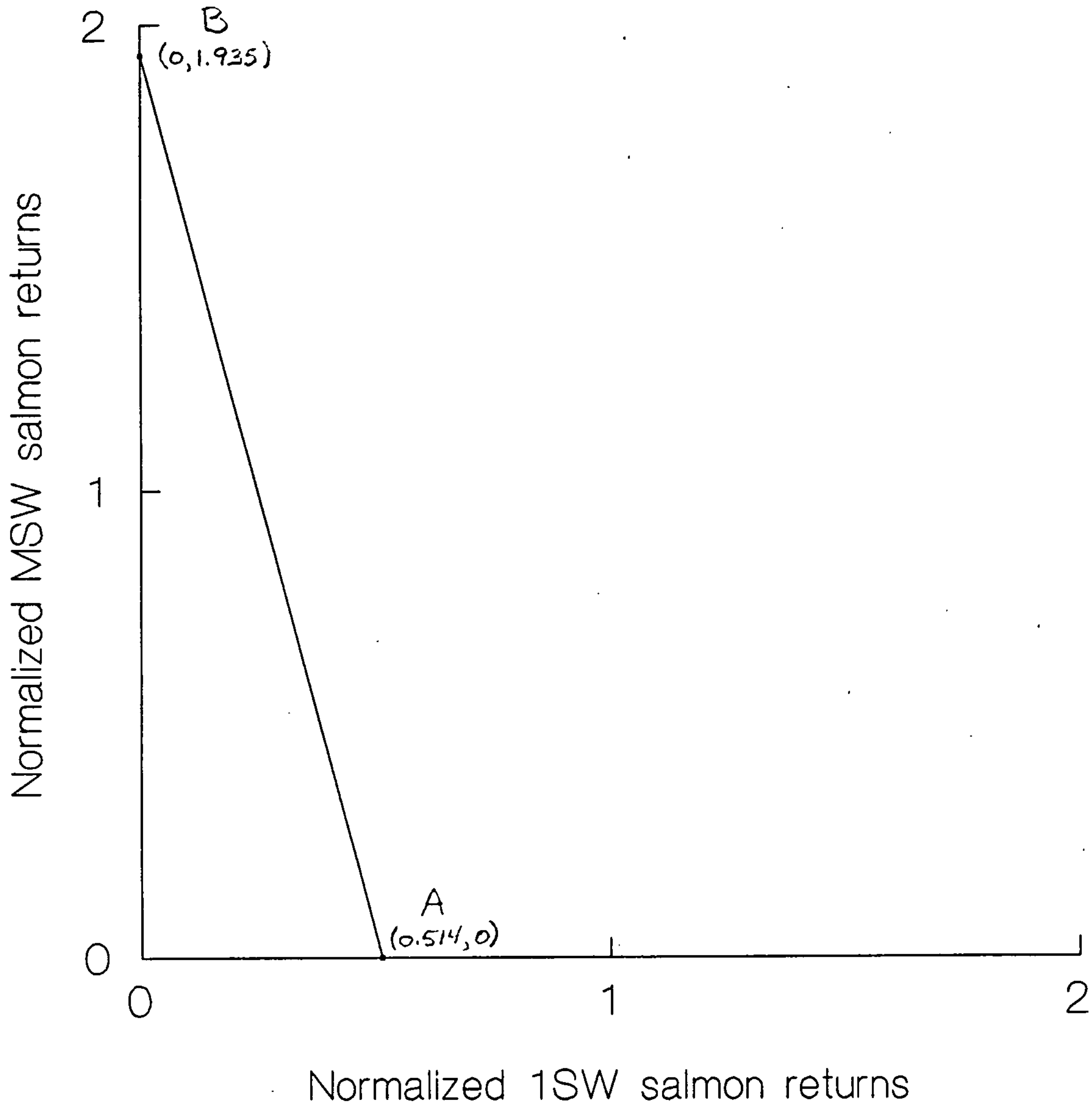

Figure 1. Graphical representation of the weighting technique for a model with two variables. The illustrated data point before normalization is (6828, 10924). This point will contribute a weight equal to the inverse of the squared distance between A and B (i.e., 1/4.008) to the likelihood function.

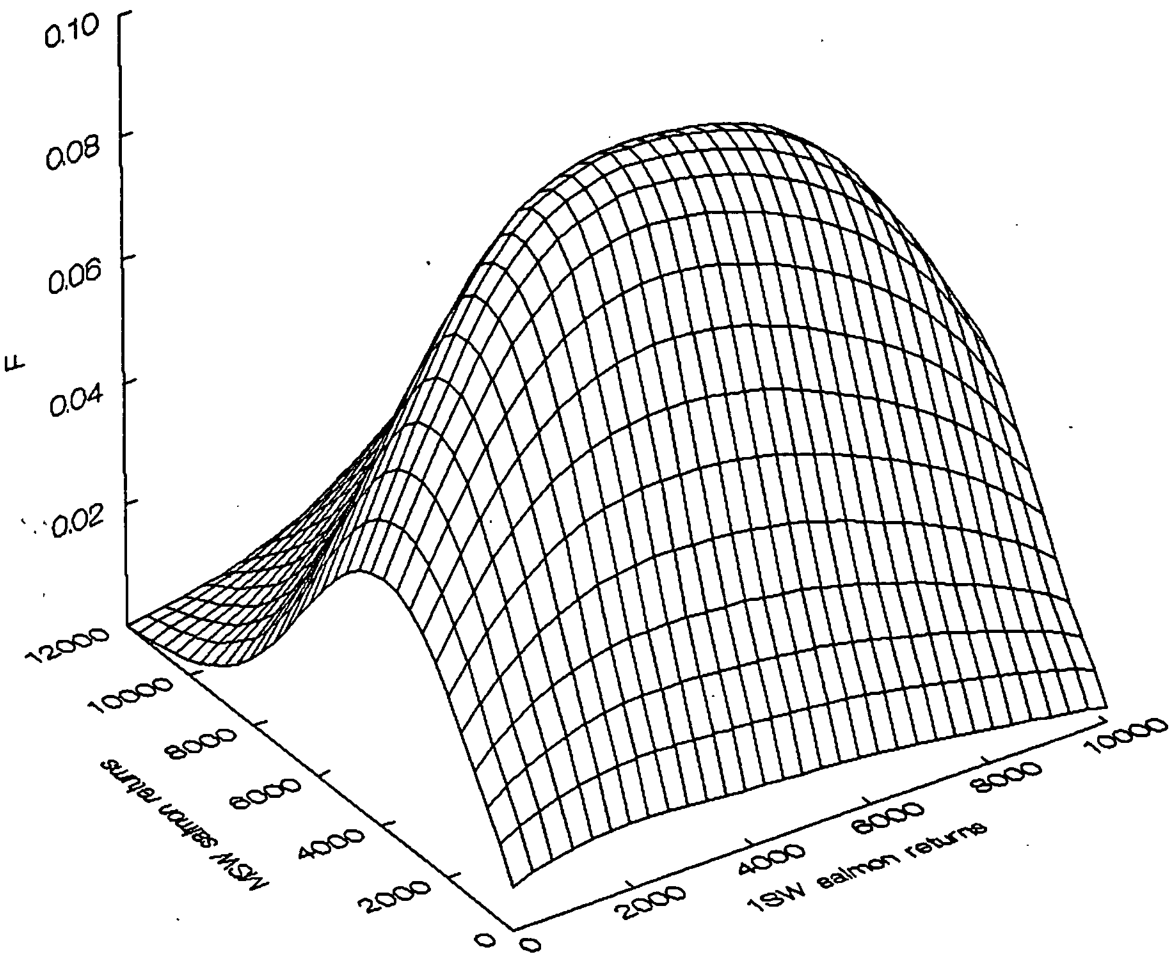

Figure 2. Probability density function for model 1 based on MSW salmon returns and 1SW salmon returns to the Saint John River.

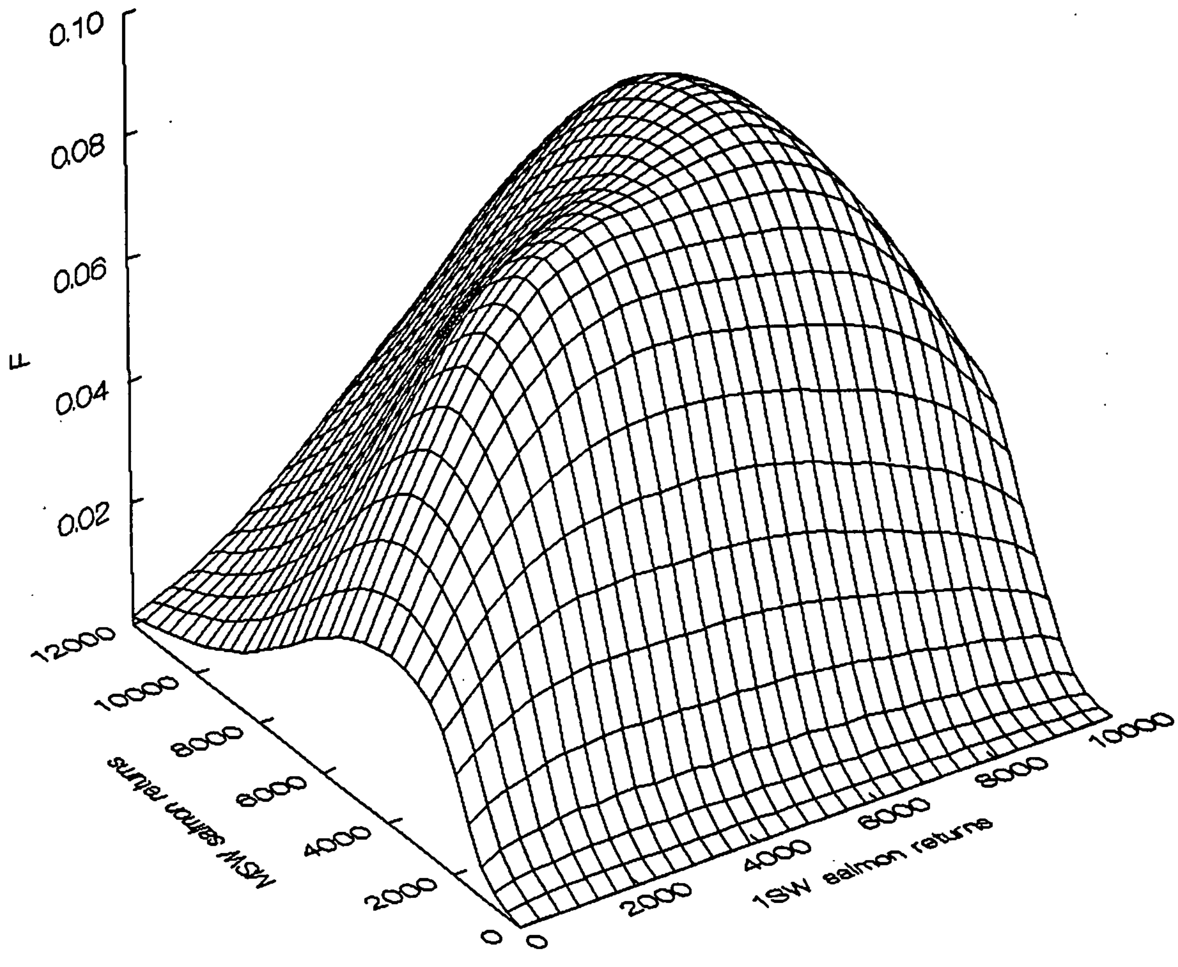

Figure 3. Probability density function for model 2 based on the natural logarithm of MSW salmon returns and 1SW salmon returns to the Saint John River.

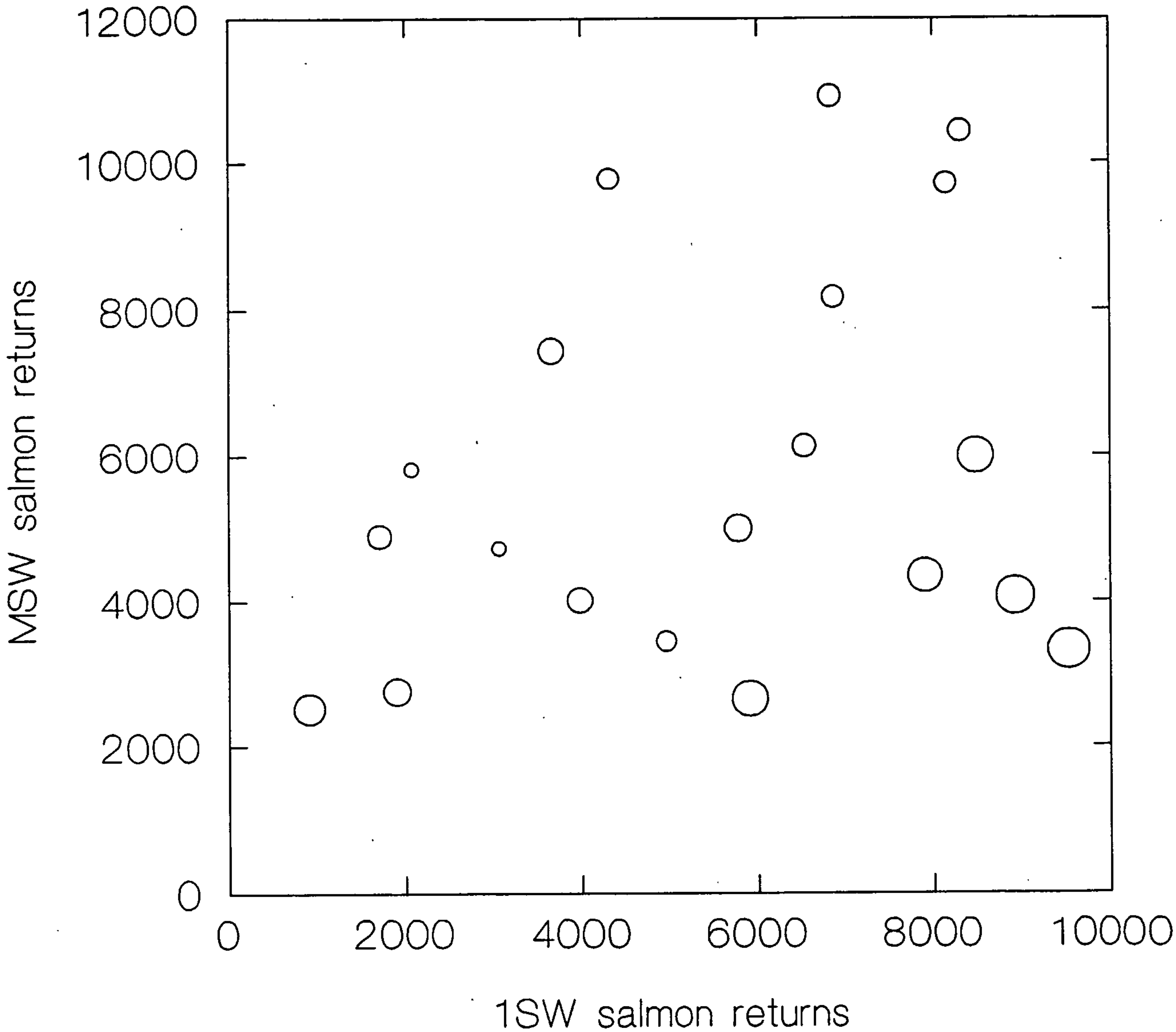

Figure 4. Plot of MSW salmon returns versus 1SW salmon returns to the Saint John River, where the size of each point is proportional to the mean fork length of 1SW salmon returns.

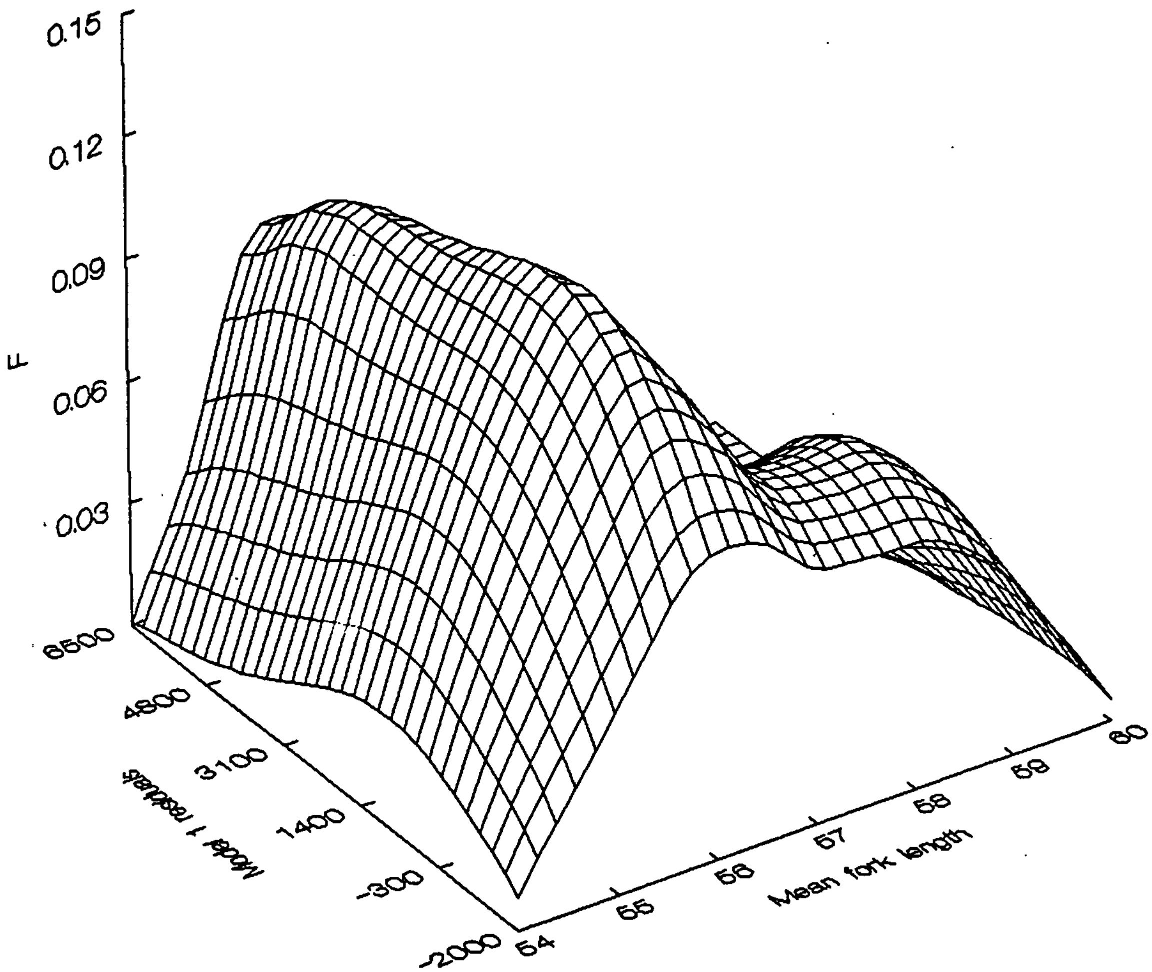

Figure 5. Probability density function, for the second step of model 3, based on the residuals from model 1 and mean fork length of 1SW salmon returns to the Saint John River.

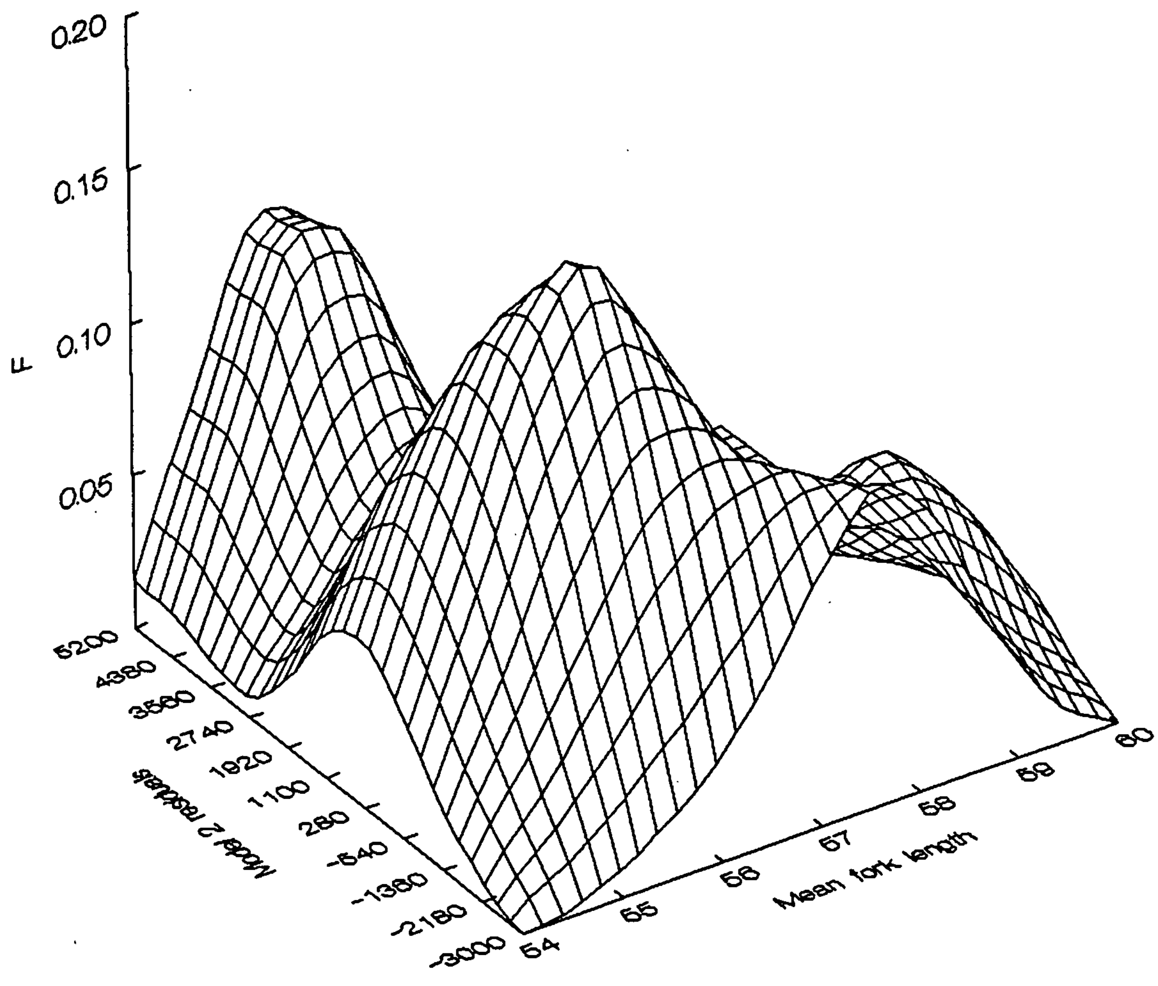

Figure 6. Probability density function, for the second step of model 4, based on the residuals from model 2 and mean fork length of 1SW salmon returns to the Saint John River.