

AN ANALYSIS OF MARK-RECAPTURE DATA

AN ANALYSIS OF MARK-RECAPTURE DATA
FROM CODED WIRE TAGGING OF HATCHERY RAISED SALMON
USING LOG-LINEAR MODELS AND GRAPHICS

By

PHILIP ERNEST JEFFREY GREEN, B.Sc.

A Project

Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Master of Science

McMaster University

September 1984

Copyright 1984, Philip Green

MASTER OF SCIENCE (1984)
(Statistics)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: An Analysis of Mark-Recapture Data from Coded Wire
Tagging of Hatchery Raised Salmon Using Log-Linear
Models and Graphics

AUTHOR: Philip Ernest Jeffrey Green, B.Sc. (University of
British Columbia)

SUPERVISOR: Professor Peter D.M. Macdonald

NUMBER OF PAGES: vi, 84

Acknowledgements

I would like to thank my supervisor, Professor Peter Macdonald, for his ability to guide me without pushing me, for his willingness to lend an ear at any moment, and for his financial support. I am grateful to Mssrs Ken Pitre and Paul Starr, from the Department of Fisheries and Oceans, for supplying the data and suggesting the project; to Mr Howard Smith and Dr Jon Schnute for helpful discussions; to the staff at Robertson Creek Hatchery for their cooperation and to Karl English from LGL Ltd. Dr Tom DiCiccio was helpful in several mathematical discussions, as were several other faculty members in the Department of Mathematical Sciences. Ms Karen Shea assisted in the coding of the graphics programs. A special note of thanks is due to my two colleagues, Ms Ellen Ng and Ms Yasasmin Sarraf, who had no idea they would hear so much about fish when they came to study statistics, and who were very encouraging. Finally, I would like to thank my wife, Dr Louise Royer, for her unending support.

ABSTRACT

In this report mark-recapture data are analyzed with the use of weighted log-linear models, mosaics, and computer drawings of fish. The data are from salmon hatcheries, subsequent returns to the hatcheries and commercial catches of salmon. The log-linear models were used to study the effects of several variables on catches and returns. It is shown that these variables may have opposite effects depending on the brood year of the fish, that hatchery returns and commercial catches do not respond in the same way to the variables, and that more research is needed to determine the causes of the brood year differences. The mosaics and fish drawings were used to study the migration of the salmon in the ocean. The results confirm that chinook salmon decrease their food intake during the return trip to the hatchery, and they are consistent with theories of ocean migration of other species of salmon.

Table of Contents

Chapter		page
I.1	Introduction	1
I.2	Review of mark-recapture techniques	8
II.1	Statistical problems with the data	15
II.2	Benefits to fishermen	16
II.3	Robertson Creek return data	25
II.4	Rosewall Creek return data	35
III.1	Migration data methods	41
III.2	Interpretation of the mosaics	52
IV	Conclusions	59
V	Recommendations for experimental design	61
Appendix		
A	Weighting the data	64
B	Description of Robertson Creek CWT data	77
C	Description of Rosewall Creek data	80
References		81

List of Figures

Figure		page
1	Location of Robertson and Rosewall Creeks	4
2a	Commercial catch regions	5
2b	Array representing catch regions	6
3	Observed and estimated returns	33
4	Troll recapture-time mosaic	45
5	Net recapture-time mosaic	46
6	Troll number-of-recaptures mosaic	47
7	Net number-of-recaptures mosaic	48
8	Fish lengths age two and three	49
9	Fish lengths age four and five	50
10-15	Effects of expansion on Poisson distribution	70

I.1 Introduction

An analysis of a large mark-recapture experiment involving chinook salmon (*Onchorynchus tshawytscha*) is presented in this report. The salmon were marked at the Robertson Creek Hatchery on Vancouver Island (Figure 1) and recovered as adults in commercial and sport fisheries from Oregon to Alaska.

The hatchery encompasses the start and end points of the chinook salmon life cycle. Fertilized eggs from the returning salmon are incubated for several months until hatching and the fry are reared in large ponds until ready for release. During rearing, some of the maturing fry are marked with a one-millimetre coded wire tag (CWT) which is implanted into the nose cartilage. The adipose fins of the tagged fish are clipped to permit their identification. The rearing fry are kept in different ponds and may be given one of several treatments. All fish treated similarly have the same tag code. When they are released, the average weight of the fish and the release date are noted. The number of tagged fish released is known from the tagging machines, the number of untagged fish is estimated from egg inventory data.

After release the fish suffer five possible fates. They may

return to the hatchery to spawn, they may be caught by troll, net or sports fishermen, or none of these. In the latter case, they die from unknown causes which may include disease and predation, usually referred to as "natural" mortality. The lifespan of these fish is not more than six years after fertilization. Throughout this report, the age of the fish refers to the number of years after the year in which the eggs were fertilized (the brood year).

Every fish returning to the hatchery is examined for a clipped adipose fin. Tagged fish are measured and their heads sent to a dissection laboratory where the tag is excised and decoded. The catches in the commercial fisheries are too large to permit examination of every fish. A proportion of the boats is sampled and the number of marks in a geographical region is inflated with an 'expansion' factor which is the reciprocal of the sampling proportion. The expanded number is reported by biweekly or monthly time intervals. The statistical effects of this expansion are considered in Appendix A. The data for the sport fishery depend on fishermen voluntarily bringing in the heads of salmon with missing adipose fins to head depots. Since the proportion doing so is unknown, the sport catches are not expanded. It is estimated that sportfishing accounts for approximately 21% of the total chinook catch (Pearse, 1982).

The data set consists of twenty tag-code groups from brood years 1972-1977. For each group the data include the number released (tagged and untagged), the pre-release treatments, average weight of the

released fish, dates of release and recapture, and hatchery return data. The recapture data are given by month for fourteen regions along the Pacific coast, ranging from Oregon State to the top of the Alaskan Panhandle (see Figure 2a), for all years in which the salmon were caught by the three types of fishing gear. In most instances, the average length of the recaptures is also reported. The hatchery return data includes only the total number of returns of each CWT group per year. There is no information available on fishing effort.

Coded-wire mark-recapture experiment are different from other mark-recapture experiments in several respects. The most important difference is that the population size is known from the outset of the experiment, whereas in most mark-recapture experiments this is the parameter that is to be estimated. The population size (N) is considered to be a known constant as it is estimated by fairly accurate volumetric counts of the eggs, although there will obviously be some error. The number of marks released (M) is exactly known. The number of marks in the fishery (m) is estimated by expanding the number in the sample. The parameter which can be estimated is what is usually referred to as the sample size (n), but what in this case corresponds to the number of salmon from Robertson Creek (or some other stock as the case may be) caught in the fishery. This is estimated by equation (1), a variation on the commonly-used Peterson estimate (Seber, 1982).

$$\hat{n} = mN/M \quad (1)$$

This technique is used by fisheries biologists to estimate the contribution of stocks of fish to a fishery.

A second aspect of this experiment which is worth highlighting is that the sampling is destructive so there is no possibility of multiple recaptures. The marking is only done at age zero, which causes some problems when estimating survival rates (Brownie et al, 1978). Finally, marked and unmarked fish return to their original point of release at the end of their life. Some of these fish spawn in the river below the hatchery; the proportion doing so is not known. These features make the data set unique with respect to most mark-recapture experiments.

The data from this experiment were used to study several important questions. The most obvious was to determine if the pre-release treatments and various covariables had any effect on the survival and catches of the fish, or on where and when they were caught. Statistical models to answer this question are presented in section II. The data also provide excellent information on the timing, direction and patterns of the migrations. This was studied using exploratory techniques in section III. The data for the net and sport fisheries were more sparse and generally of lower quality than the troll fishery, which is a directed, as opposed to an incidental, fishery. In section II only troll fishery data were used. The net data had no Alaska recaptures.

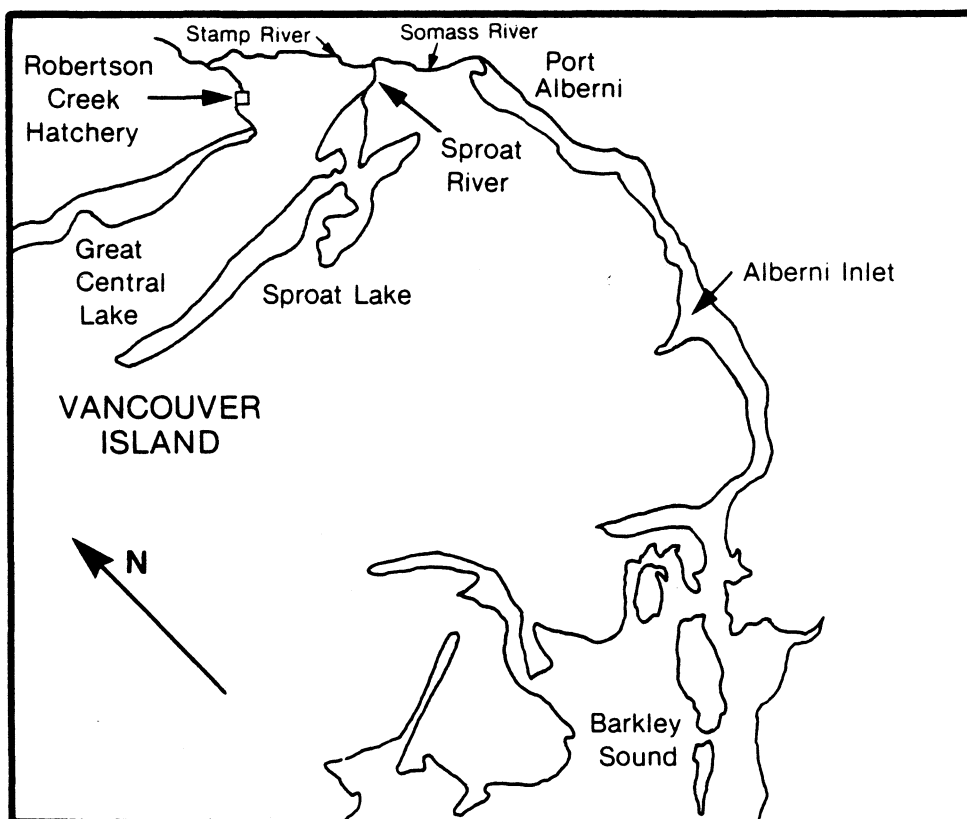
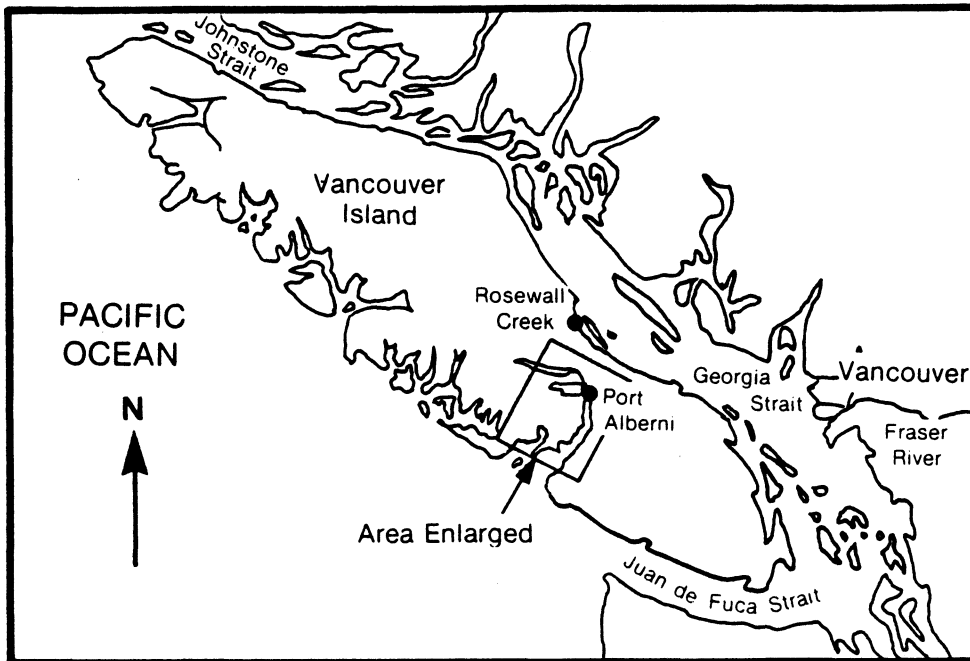


Figure 1. Location of Robertson and Rosewall Creeks.

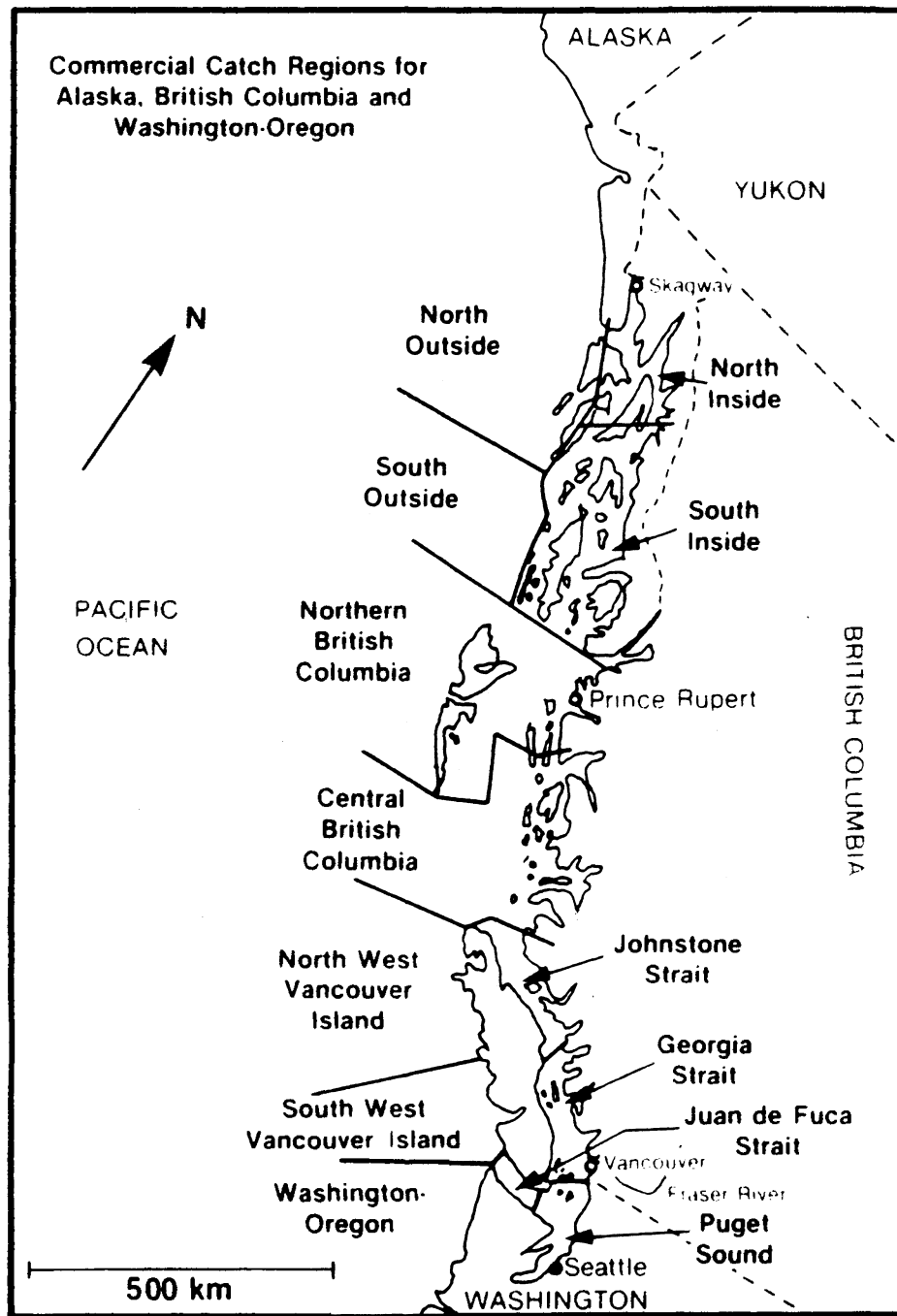


Figure 2a.

Array Representing Commercial Catch Regions

North Outside	North Inside
South Outside	South Inside
Northern British Columbia	Alaska Unknown
Central British Columbia	British Columbia Unknown
North West Vancouver Island	Johnstone Strait
South West Vancouver Island	Georgia Strait
Washington- Oregon and Puget Sound	Juan de Fuca Strait

Figure 2b.

I.2 Review of mark-recapture techniques for commercially exploited populations.

Seber(1982) gives an excellent review of mark-recapture models for commercially exploited populations. The basic model rests on the assumption that the population of fish decreases in an exponential fashion, with two instantaneous mortality rates, fishing (F) and natural mortality (M). Thus

$$E(N_t) = N_0 \exp(-(F+M)t) \quad (2)$$

where N_0 is the number of marks at time 0, and N_t is the number of marks at the start of the t^{th} interval. The unit time interval is considered to be one year's duration.

Variations on this model have been developed which assume that the coefficients of mortality are either constant or a function of time. These are described by Seber (1982). Gulland (1955) derived maximum likelihood estimates for F and M. He assumed both coefficients were constant, that the time to recapture was exponentially distributed with parameter $Z=F+M$, and that the total number of

recaptured fish at the end of the experiment was binomially distributed. Nicholson and Pope (1977) applied these methods to a population of herring, and used a refinement of Gulland's techniques due to Chapman (1961).

In this study it is not possible to assume that either F or M are constant with respect to time or age. It is clear that the probability of being caught is age dependent. The 'plugs' or lures used in the troll fishery are about fifteen centimetres long, which is comparable in size to a juvenile salmon. The fish must grow to a large enough size before they are catchable (often called recruitment of fish to the catchable population). The number of recaptures increases with age before dropping off, and as we know that the population is not increasing we must conclude that fishing mortality is not constant. Biological populations are usually found to have higher mortalities at the earlier stages of life, and there is no reason to believe differently in this case, especially considering that marking itself may cause some mortality. There may be some grounds for assuming that natural mortality is constant in the adult population.

In order to understand the process more clearly, let us for the moment consider a model with only one type of mortality, Z_i , which is dependent on the age i . Thus $\exp(-Z_i)$ is the probability of surviving the i th interval. Equation (2) can then be rewritten:

$$E(N_t) = N_0 \exp\left(-\sum_{i=0}^{t-1} Z_i\right) \quad (3)$$

The development will be clearer if we rewrite

$$q_i = \exp(-Z_i)$$

thus,

$$E(N_t) = N_0 q_1 q_2 \dots q_{t-1} \quad (4)$$

The expected number dying (from all causes) in the t^{th} interval is:

$$E(D_t) = N_0 q_1 q_2 \dots q_{t-1} (1 - q_t) \quad (5)$$

In an animal population where all individuals died by their fourth year, for example, the probabilities of dying each year could be described as follows:

$$\begin{aligned} \text{Prob}(\text{die in 1st interval}) &= 1 - q_1 \\ \text{Prob}(\text{die in 2nd interval}) &= q_1(1 - q_2) \\ \text{Prob}(\text{die in 3rd interval}) &= q_1 q_2(1 - q_3) \\ \text{Prob}(\text{die after 3rd interval}) &= q_1 q_2 q_3 \end{aligned} \quad (6)$$

Which could be reparameterized by

$$\begin{aligned}
1-q_1 &= \theta_1 \\
q_1(1-q_2) &= \theta_2 \\
q_1q_2(1-q_3) &= \theta_3
\end{aligned}
\tag{7}$$

Since $D_4 = N_0 - D_1 - D_2 - D_3$ and N_0 is known, then

$$\theta_4 = 1 - \theta_1 - \theta_2 - \theta_3 \tag{8}$$

and we need only estimate three parameters. Obviously, the constraint

$$0 \leq \theta_i \leq 1$$

must hold. As stated, the mortalities are a function of age and are also a function of a vector of covariables $\underline{Z}$. Thus

$$\theta_i = \text{fn}(\text{Age}_i, \underline{Z}\beta) \tag{9}$$

The function on the right hand side of (9) could be described by a general linear model. Multiplicative effects seem most plausible. A model of the form

$$\theta_i = \exp(\alpha_i + \underline{Z}\beta) \tag{10}$$

is proposed. Now let us consider the situation with two competing causes of death (often called the competing risks problem (Kalbfleisch and Prentice, 1980)), fishing and natural mortality. In the continuous case, the instantaneous coefficients of mortality are usually called hazard functions in survival data analysis. The subdensity function (Kalbfleisch and Prentice, 1980) for fishing mortality is

$$f_F(t) = F(t) \exp\left(-\int_0^t F(u) + M(u) du\right) \quad (11)$$

where $F(t)$ and $M(t)$ are the hazards due to fishing and natural mortality at time t . The vector of covariables is omitted for simplicity. The probability of dying from fishing and doing so before time t is

$$\text{Prob}[T < t, \text{death by fishing}] = \int_0^t f_F(u) du \quad (12)$$

where T is the lifetime. In the case where $F(t)$ and $M(t)$ are constant,

$$\text{Prob}(T < t, \text{death by fishing}) = (F/(F+M)) (1 - \exp(-(F+M)t)) \quad (13)$$

which is the same as formula (1) of Gulland (1955). If $F(t)$ and $M(t)$ are constant over each interval, then the marginal density function for fishing mortality over interval i is

$$f_{Fi}(t) = F_i \exp(-(Z_1 + \dots + Z_{i-1} + Z_i(t-(i-1)))) \quad (14)$$

where $i-1 \leq t < i$ and $Z_i = F_i + M_i$

The probability of death by fishing in the interval $[i-1, i)$ is:

$$\begin{aligned} \int_{i-1}^i f_{Fi}(t) dt &= F_i \exp(-\sum_{j=1}^{i-1} Z_j) \int_{i-1}^i \exp(-Z_i[t-(i-1)]) dt \\ &= F_i / Z_i \exp(-\sum_{j=1}^{i-1} Z_j) (1 - \exp(-Z_i)) \end{aligned} \quad (15)$$

Returning to the notation of (5), we have

$$E(C_t) = (F_t / (F_t + M_t)) N_0 q_1 q_2 \dots (1 - q_t) \quad (16)$$

where C_t is the number caught by fishing in interval t , that is, the fishing mortality.

In the data set under study there are no data on natural mortality. The total natural mortality can be estimated but as the times of death are not known it is not possible to estimate the M_i . There is no way to estimate all the parameters in (16) unless prior M_i are assumed. As Lawless (1982, p491) points out, estimating a hazard function that would result if other causes of death were removed requires strong assumptions. To assume the M_i (i.e. a discrete hazard function for natural mortality) requires the researcher to be very clear about the context in which they are being assumed, with or without fishing mortality. There may be mortalities due to fishing caused by damage to fish that escape or predation by seals or sealions while on the hook (Ricker, 1976) which are not observed as fishing

mortality and would be incorrectly classified as natural mortalities. Thus natural mortality could be different at different fishing mortalities. To avoid assuming the M_i we will again reparameterize:

$$\theta_i = (F_i / (F_i + M_i)) q_1 q_2 \dots q_{i-1} (1 - q_i) \quad (17)$$

The θ_i are the probabilities of recapture. We can then set up a linear model of the same form as (10). Since there are five causes of death and we are modelling the probabilities of recapture by the troll fishery only, the M_i can be considered to be the sum of the other four types of mortality.

In addition to fishing mortalities being age dependent, initial rough plots of recaptures at each age, location and brood year showed that they are also dependent on both location and brood year. Thus the model should include location and brood year as effects. This means the q_i will be specific to the location modelled. Fishing mortality is clearly dependent on fishing effort, for which there are no data. To some extent, differences in fishing effort from year to year will be compensated for by the brood year effect included in the model.

Equation (10) is essentially a type of log-linear model. Log-linear models have previously been used for mark-recapture data by Fienberg (1972) and Cormack (1981). Both of these applications involved multiple recaptures.

Part II Analyses of Responses to Treatments

II.1 Statistical Problems with the Data

The multiplication of the recaptures found in the sampled boats by an expansion factor causes some problems for the analysis of the data. First there is an inflation of the variance, since

$$\text{Var}(kX) = k^2 \text{Var}(X)$$

Second, the shape of the distribution of expanded recaptures will differ from that of the true recaptures.

The first problem can be tackled by appropriately weighting the data. To see how this should be done, refer to appendix A. The effects on the shapes of the distributions are also discussed in Appendix A.

II.2 Benefits to Canadian and Alaskan fishermen.

The Robertson Creek Hatchery and its operations are funded by the federal government of Canada. The fish produced by the hatchery are caught by both American and Canadian fishermen, most of the American catch being in Alaska. The managers of the hatchery expressed an interest in knowing whether their efforts were resulting in increased incomes for Canadian or American fishermen (or both) and whether there were any actions they could take to direct any increase to Canadian fishermen. These actions would be in the form of modifying release weights, days of release, or any pre-release treatment. The model did not include any parameters for dollar income; the modelling effort was an attempt to model the response (number of recaptures) in Alaska and British Columbia as a function of various effects and covariables. It could therefore be used not only to answer the hatchery managers' questions, but also to study in general the factors causing catches to increase or decrease. The model was of the form:

$$E(Y) = N \exp(\alpha + Z\beta) \quad (18)$$

where α is a vector of treatment effects, Z is a vector of covariables, and β is a vector of regression coefficients.

The factors or effects modelled were:

- age at recapture (AGE)
- pre-release treatment (TMT)
- brood year (BYEAR)
- place of recapture (PLACE)

The covariables were:

- weight at release (WT)
- Julian day of release (JDAY)

The model was fit using the computer program GLIM on a CDC 730 computer (as were all models in section II).

Before fitting was begun some preliminary plots and calculations were done. The calculations showed that the ratio of recaptures in Alaska to those in B.C. varied considerably over the lifetimes of the fish (see table 1). At age two there were no recaptures in Alaska and very few in B.C. At age three there were generally more caught in B.C. than Alaska but at ages four and five this ratio was reversed. At age six there were hardly any fish caught. The fish at age two are small, have a low commercial value, and are not caught in large quantities. At age six they have a higher commercial value but are also not numerous. For these reasons I decided to model only ages three to five.

Plots of the logarithms of the expanded recaptures at the three ages and two locations against weight at release and Julian day showed evidence of linear trends but involved two CWT groups which were

clearly outliers. Their release weights were outside the range of the other groups, so they were not included in the model. A weight vector was defined corresponding to the sampling fractions. The sampling fractions for yearly totals were known for most of the CWT groups at the three ages and two locations. Where they were not known the mean sampling fraction was used.

The emphasis during model fitting was not so much to find a "best" model but rather to determine which sources contributed most to a reduction in deviance. Since this is a "messy" data set (in the sense of, for example, Urquhart and Weeks, 1978), fitting high-order interactions leads to much aliasing and results difficult to interpret. Fitting only main effects and first order interactions leads to a more parsimonious model.

An analysis of deviance (anodev) is presented in table 2. With discrete data and large counts, such as these data, the deviance may be approximated by chi-square (McCullagh and Nelder, 1983,p78). However in this data set, even with weighting, there are distributional uncertainties and there is probably some over-dispersion, and I am not sure how well the approximation holds. The importance of an effect or covariable was deduced by its relative reduction in deviance.

Table 1

Alaska/British Columbia Recapture Ratios by Age
(Troll Fishery)

Group	Age 2	Age 3	Age 4	Age 5	Age 6
1		.31	BC	5.14	
2		BC	1.55	2.23	AL
3		.35	.67	3.33	AL
4		.05	.83	1.67	
5		BC	.93	.67	
6		.12	2.00	AL	
7		.11	3.38	1.77	
8		BC	.44	.60	.67
9	BC	.49	1.23	2.95	
10		.53	1.13	3.92	
11		.39	.84	AL	
12	BC	1.25	1.67	2.00	
13	BC	.28	1.14	1.15	
14	BC	.19	1.27	2.18	
15		.24	3.00	2.00	
16	BC	.35	.99	3.86	
17	BC	.26	1.04	1.23	
18	BC	1.15	1.79	1.79	
19		.09	1.80	2.86	
20		.63	1.79	5.33	

BC: fish only caught in British Columbia
AL: fish only caught in Alaska

Table 2

Analysis of Deviance (Anodev) for the Recapture Model (Section II.2)

Model	Source	Deviance	d.f.	Dispersion
	Mean	3223	107	30.12
A	+AGE+PLACE	2147	104	20.64
B	A+BYEAR	1096	99	11.07
	B+TMTS1	1055	97	10.88
C	B+TMTS2	767	95	8.07
D	C+AGE.PLACE	462	93	4.97
	D+AGE.PLACE.JDAY	365	87	4.19
E	+AGE.PLACE.WT	314	81	3.88
	D+AGE.PLACE.WT	378	87	4.35
E	+AGE.PLACE.JDAY	314	81	3.88
	E+BYEAR.AGE	182	71	2.56
	+TMTS2.AGE	172	63	2.73
F	+BYEAR.AGE.PLACE	111	48	2.32
G	E+BYEAR.PLACE	286	76	3.76
H	E+BYEAR.AGE	182	71	2.56
I	E+TMTS2.AGE	239	71	3.36
J	H+BYEAR.AGE.PLACE	119	56	2.13

Results of the model.

Model A is the simplest model presented and only includes main effects. The reduction in deviance achieved by adding BYEAR shows that there are very large differences in the recapture rate for different brood years. These differences are greater than those due to the different treatments.

There are two sets of treatments (see Appendix B), TMTS1 and TMTS2. The first set includes CWT groups two and three as production groups. The reduction in deviance obtained by including TMTS2 is much larger than that obtained using TMTS1, so the food pellets these two CWT groups were given had some effect. CWT group two showed a greater recapture rate than group three, which was less but quite close to the production groups. The most successful treatment was treatment six (mid release).

Model D shows that there is a large age-place interaction. Thus the recapture rate at different ages varies from Alaska to B.C. more than one would expect by considering the average recapture rates in these two locations alone.

Model E fits individual regression coefficients for recapture rates as a function of release weight and Julian day at each age and place. The two routes taken to arrive at model E show that Julian day of release has a stronger effect on recapture rates than weight at

release. All the slopes for Julian day are negative when it is entered first, except for age five in B.C. which is positive. This suggests that as Julian day of release is increased, recaptures in B.C. increase while in Alaska they decrease. When the release weights are added this difference in slopes disappears; however a similar one appears for the release weights, indicating that as release weights rise, recaptures in B.C. increase slightly faster than in Alaska. Note that this is only in the range of modelled weights and Julian days, and the relations will not necessarily hold true when extrapolated.

Models F-J examine some of the interactions. It appears from these models that the brood year-age interaction is important and the others much less so.

Discussion.

I believe the most important result from these models is the large reduction in deviance achieved by adding the brood year effect. This could mean that there are large yearly changes in fishing effort; in the genetics of the broods, in the conditions at the hatchery, such as water quality or water level in the rivers leading to the ocean, in mortality after marking and other factors.

The differences in recapture rates in Alaska and B.C. due to release weight and Julian day are slight and I doubt whether it is

worth trying to manipulate these in an attempt to increase recaptures in B.C. It is difficult to interpret the increasing recaptures due to increasing Julian day in the light of the fact that mid release (treatment six) was found to increase recaptures more than early or late release (treatments five and seven).

The large deviances in the anODEV table can be misleading. The deviance is really a measure of discrepancy between the fitted and observed values. With large sample sizes this discrepancy is easier to detect, even though it may be acceptably small. What is needed is a measure of whether the discrepancy between observed and expected values exceeds an acceptable amount. An illustration of this problem is provided in the next section.

The model presented here, or other similar models, could be used to predict the catch of stocks of salmon in different locations. The model could be adapted to fit the requirements of fishery managers; for example, they may wish to have predicted catches broken down into smaller geographical regions. Confidence intervals could also be constructed for the predictions, as is done in the next section. Since it is not possible to know what the brood year effect is until after the entire brood is dead, or until we learn the causes of the brood year effect, the manager could be asked to select between optimistic and pessimistic scenarios, and brood year effects would be chosen based on previous estimates. This model would be more credible for predicting catches than a simulation model not based on parameters estimated from data. Of course, the dangers of extrapolation would

have to be born in mind.

II.3 Hatchery Return Data.

In order for the hatchery to increase the number of salmon at sea and available for the fisheries, the hatchery managers must know the factors which affect the probabilities of capture. Are these the same factors that influence the number of salmon returning to the hatchery? By analyzing the hatchery return data we can attempt to answer this question.

When the migrating salmon return to the hatchery at maturity, all of them are sampled. There are no expansion factors, no variable fishing effort, and so the statistical problems encountered in the previous section are greatly simplified. The major problem encountered in analyzing these data is that not all of the fish that return to spawn return to the hatchery: many spawn in the river below the hatchery, and nothing is known about the proportion doing so. Factors affecting the number spawning below the hatchery could include water conditions and crowding at the hatchery entrance.

The factors which we may wish to include in a model of these data are the ages (AGE) of the returning fish, their brood year (BYEAR), the year they returned in (RYEAR) and the prerelease treatment (TMT). The effect of the average release weight and Julian day of release

should also be included (WT and JDAY). Obviously, AGE and BYEAR will be aliased with RYEAR since $\text{return year} = \text{age} + \text{brood year}$. This means there is no way of estimating effects due to water conditions in the year that the fish return separately from age and brood year effects.

The data set consists of four age classes for the same twenty CWT groups, two of which are anomalous, giving a total of seventy two data points. The data space is quickly exceeded by the parameter space when the interactions are fit (well before the saturated model is fit) and there is much aliasing, especially between brood year and treatments. When the brood year X treatment interaction is fit, there are thirty parameters, twenty-eight of which are aliased. This forces us to look for a very parsimonious model just in order to obtain a fit at all. With four ages, six brood years and seven treatments, there are a total of one hundred and sixty-eight cells. Less than seventy-two of these have data in them since some have replicate observations. With such messy data, we should not be surprised if the fit is not particularly good. The emphasis should be on trying to find effects which contribute a relatively large amount to the reduction in deviance, as in the previous section.

An analysis of deviance is presented in table 3. The models are all of the same form discussed earlier: we are modelling the probability of a fish returning to the hatchery. The treatments are the same as the second group of treatments used in the recapture model.

Table 3

Analysis of Deviance (Anodev) for the Hatchery Model (Section II.3)

Model	Source	Deviance	d.f.	Dispersion
	Mean	5115	71	72.04
	+AGE	3309	68	48.66
	+BYEAR	1349	63	21.41
A	+TMT	1113	59	18.86
A1	A+WT	1081	58	18.64
A2	A+JDAY	994	58	17.14
B1	A+AGE.WT	929	55	16.89
B2	A+AGE.JDAY	851	55	15.47
B3	A+BYEAR.WT	1007	54	18.65
B4	A+BYEAR.JDAY	983	55	17.86
B5	A+TMT.WT	1024	56	18.29
B6	A+TMT.JDAY	980	56	17.50
C	A+AGE.BYEAR	570	43	13.26
C1	C+AGE.BYEAR.WT	362	21	17.26
C2	C+AGE.BYEAR.JDAY	218	20	10.90
D	A+AGE.TMT	837	43	19.47
D1	D+AGE.TMT.WT	515	29	17.76
D2	D+AGE.TMT.WT	532	28	19.00
E	A+BYEAR.TMT	1024	56	18.29
E1	A+BYEAR.TMT.WT	2851	59	48.32
E2	A+BYEAR.TMT.JDAY	2786	59	47.22

Results of model and discussion.

Model A includes main effects only, all of which significantly reduce the deviance. Models A1 and A2 show that Julian day is the most important of the two covariables when all response categories are given the same slope (with respect to the covariables).

Models B1-B6 examine reductions in deviance obtained by fitting different slopes for the two covariables at each of the three main effects. None of these models are clear "winners". Julian day causes a greater reduction in deviance than weight in all cases, especially in interaction with age.

Models C, D and F fit first-order interactions between main effects and a different slope for JDAY and WT at each of these interactions, that is, at all cells with the same levels of the two components of the first-order interaction. It is not possible to fit any more interactions than this due to the dimension of the data space. While in some cases there may be enough degrees of freedom remaining to fit an interaction, during the iterative fitting process the deviance begins to increase and GLIM cannot fit them. The model with the lowest dispersion and deviance is model C2. Model C1 is also quite good. These two models show that including weight or Julian day

with the age-brood year interaction significantly improves the fit. The hatchery managers are well aware that weight and time of release are very important in determining the number that return. These models show that weight and time of release have different effects depending on the age and brood year of the fish that return. Comparing these effects between age-brood year combinations may show little difference, but accounting for all of these differences improves the models overall ability to predict returns.

Again the modelling effort shows that differences between brood years are very significant and emphasizes the need for further research on the causes of these differences. Bilton et al. (1982) analyzed the influence of time and size at release of juvenile coho salmon on returns at maturity to Rosewall Creek (see Figure 1) by using a response surface analysis. However the entire experiment was carried out for only one brood. While a certain combination of size and time at release may produce high returns within a given brood, the differences in returns between broods might completely overshadow any gains obtained by manipulating hatchery conditions. The fact that brood year is such a strong effect in both hatchery return data and the recapture data shows that if fishing effort causes the effect, then fishing must kill a high enough proportion of the fish that there are significantly fewer fish available for spawning in the hatchery. However, if the effort in the troll fishery causes the observed brood year effect in the recapture models, the terminal fishery (which operates in Alberni Inlet just before the salmon migrate up the Somass

River to Robertson Creek) could cause the brood year effect in the hatchery return models (see Schnute and Sibert, 1983). Terminal fishery catches should be included in this model. There are also sport and Indian fisheries but no data are available for these. It seems to me that a matter of utmost importance is to collect some data on the effort in each fishery. Without these data it is impossible to determine whether there are factors other than fishing which influence both recaptures and returns, and, given the strength of the brood year effect, this is a serious gap in our understanding of the resource.

Table 4 shows the sign of the slopes of the returns with respect to the two covariables at different brood years. The striking feature of this table is that in brood year one (1972) the estimates of the slopes are all greater than or equal to zero, whereas in brood year four (1975) they are all less than or equal to zero. If two researchers had done separate mark-recapture experiments in these years, they would have reached opposite conclusions! This throws into question the value of the experiment of Bilton et al. (1982). Their experiment should be repeated in other years.

Model C2 was used to demonstrate the use of this type of model for predicting returns. In Figure 3 the observed and predicted returns are given with approximate 95% confidence intervals. This model (or a similar one) could be used to determine if there was a differential mortality between marked and unmarked fish. The model estimates the percentage of returning MARKED fish. It could easily be used to predict the percentage of returning unmarked fish. By making

these predictions for several years' releases the total number of returning fish to the hatchery each year could be estimated. If these numbers were consistently lower than the observed numbers, this would be an indication of greater mortality for marked fish.

In Figure 3 the bars represent the observed percentage return to the hatchery of the CWT groups. The centre of the circles are the predicted percentage returns, and the two lines represent approximate 95% confidence limits for the predictions. These two lines are often so close together that they appear as one line. The triangles indicate non-fitted values. It is apparent that the confidence limits do not always include the observed number of returning fish, and this is a weakness in this analysis. The critic must bear in mind however, that the difference may only be a small fraction of a percentage point, which for practical purposes, and given the messiness of the data, is quite tolerable.

If data were available on the sex of the returning fish the fit could be improved, since males and females do not return to the hatchery at the same ages in the same proportions. There are virtually no two or three year old females returning to the hatchery, and no five year old males.

Table 4

Slopes of Responses to Covariables Weight (w)
and Julian day (j) for Hatchery Return Model

Broodyear	Age 2		Age 3		Age 4		Age 5	
	w	j	w	j	w	j	w	j
1	A	+	+	0	+	0	+	A
2	-	-	0	-	0	0	0	-
3	0	0	0	-	0	0	0	0
4	-	0	-	0	-	0	-	0
5	0	0	0	0	+	0	0	0
6	0	0	0	0	0	0	A	0

+ Positive slope

- Negative slope

0 Not significantly different from zero

A Aliased, cannot be estimated

OBSERVED AND ESTIMATED PERCENTAGE RETURN

1.2X-

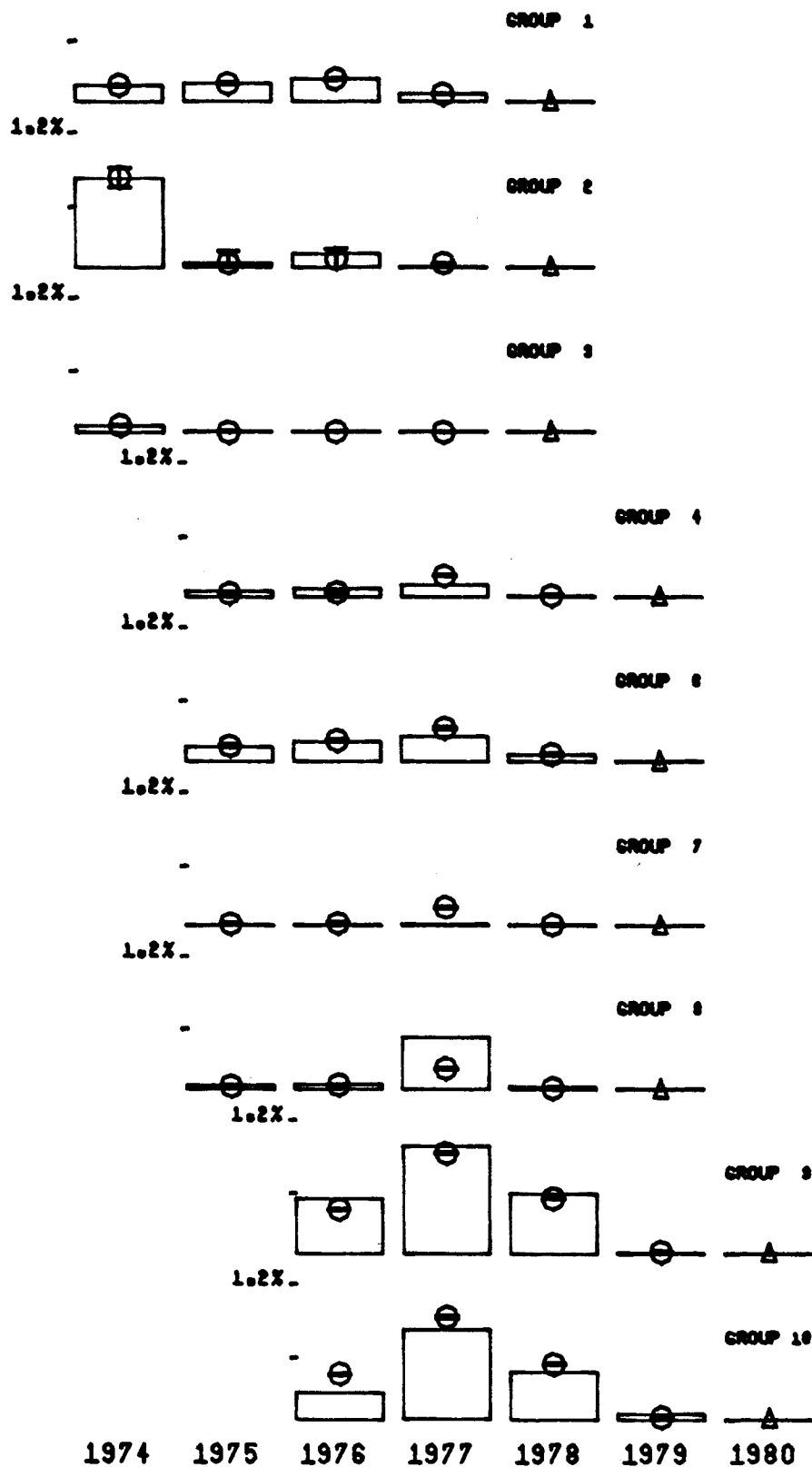


Figure 3.

OBSERVED AND ESTIMATED PERCENTAGE RETURN

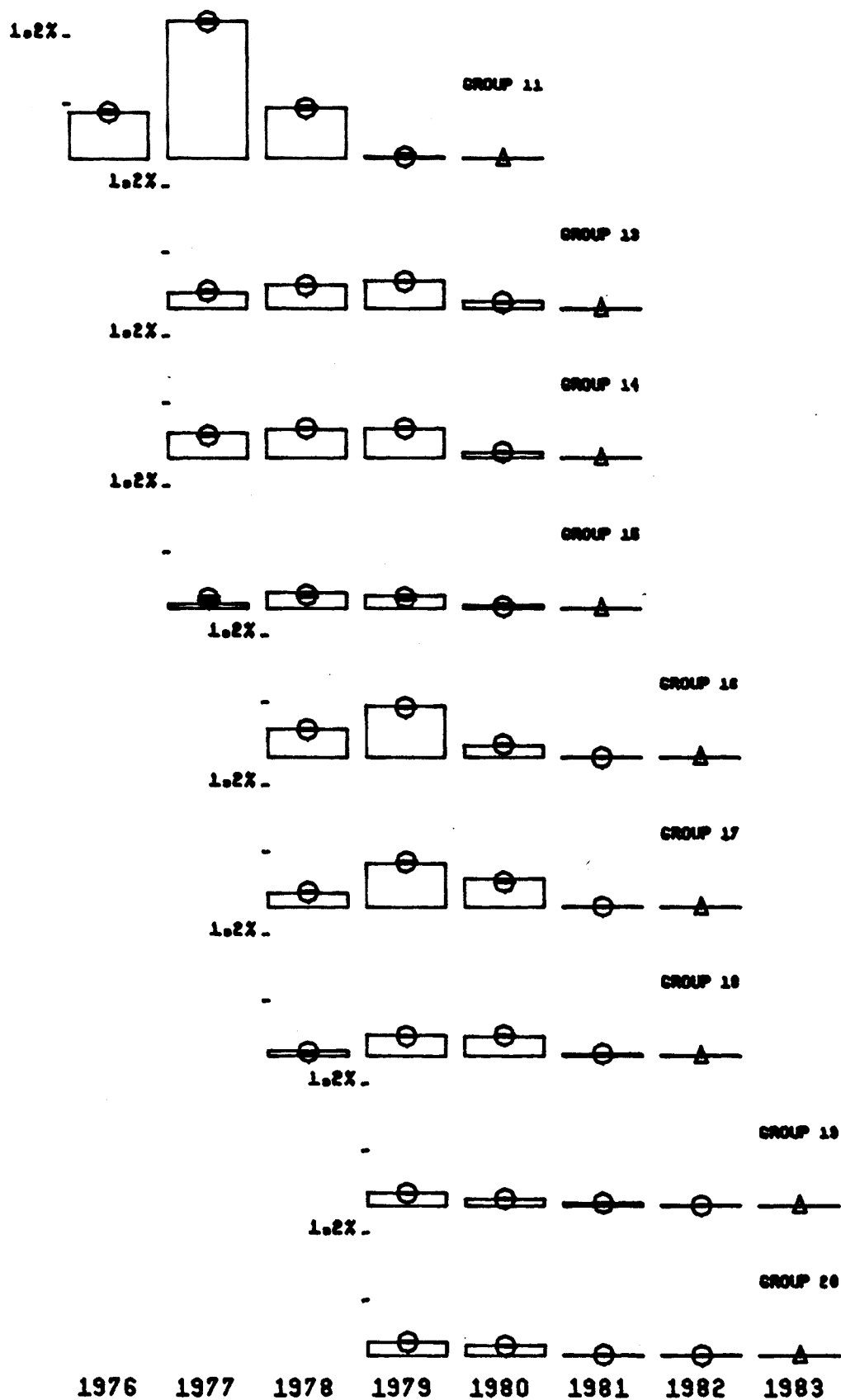


Figure 3, continued.

II.4 Analysis of the Rosewall Creek Data.

As mentioned in the previous section, Bilton et al. (1982) used a response surface analysis to analyze the influence of time and size at release of coho salmon on their returns at maturity to Rosewall Creek (Figure 1). The experiment was essentially a factorial design (see Appendix C for details). They analyzed the sum of the returns and commercial and sport catches of adult coho salmon, but only hatchery return data for jacks (males that mature early). They assumed that the effects of time and size at release were additive and that errors were normally distributed and additive. Their model was non-linear and consisted of eight unknown parameters which had no simple biological interpretation.

I felt that the type of model presented so far in this report would be more suitable for analysis of this set of data. One problem with the model of Bilton et al. (1982) was the assumption of normal errors. As the authors themselves admit, with this assumption the predicted responses could take negative values, which is of course physically impossible. The effects were also entered additively, even though this is rarely done with count data (Bishop et al., 1975,

pp23-4, McCullagh and Nelder, 1983). There were four release dates, they treated Julian day of release as a continuous covariable; whereas I decided to treat it as a four level factor since the data were already grouped. This has advantages in estimating interactions.

Adding the hatchery returns to the commercial and sport catches could have implications for the variance structure of the data. As discussed in section II.1, the variance and distributions of the commercial catches are affected by the expansion factors used, whereas the hatchery returns are not. How should the data be weighted when part of the count is expanded and part is not? There is a more fundamental reason for not adding the two counts: do the catches and returns really respond in the same way or should separate modelling attempts be made as in sections II.2 and II.3 of this report?

The following factors were included in the models:

- pond (POND)
- Julian day (JDAY)
- Age (AGE)

The only covariable was release weight (WT). Only the hatchery returns were modelled as there were no data available on the expansion factors. An analysis of deviance is presented in table 5.

Table 5

Analysis of Deviance (Anodev) for the Rosewall Creek Model
(Section II.4)

Model	Source	Deviance	d.f.	Dispersion
	Mean	3703	111	33.36
A1	Mean+AGE	3286	110	29.87
A2	Mean+JDAY	2076	108	19.22
	Mean+POND	3639	106	34.33
A3	A1+JDAY	1659	107	15.50
	A1+POND	3223	105	30.70
	A2+POND	2026	103	19.67
A4	Mean+AGE+POND+JDAY	1610	102	15.78
A5	A4+WT	1412	101	13.98
B1	A5+AGE.JDAY	1246	98	12.71
	A5+AGE.POND	1231	96	12.82
	A5+JDAY.POND	1348	91	14.81
	A5+AGE.WT	941	100	9.41
	A5+JDAY.WT	1354	98	13.82
	A5+POND.WT	1408	96	14.67
C	A5+AGE.JDAY.WT	371	94	3.95
	A5+AGE.POND.WT	895	90	9.94
	A5+JDAY.POND.WT	1313	83	15.82
D1	A3+AGE.JDAY+AGE.JDAY.WT	333	96	3.47
D2	D1+POND	311	91	3.42

Results and discussion of model.

The main effects AGE and JDAY were significant irrespective of their order of entry into the model. For practical purposes it is not really worth including the POND effect: the six different populations contribute relatively little to reducing the deviance, although in a strictly statistical sense the reduction is significant, assuming the deviance is distributed as a chi-square random variable.

Amongst the first-order interactions involving only the factors, AGE.JDAY is the strongest, amongst those involving the covariable weight, AGE.WT is the strongest. However, adding the second-order interaction AGE.JDAY.WT to the model drastically reduces the deviance. Models D1 and D2 are proposed. Essentially the AGE.JDAY terms are intercepts and the AGE.JDAY.WT terms are slopes for the log of the returns as a function of weight at each age-Julian day combination. Plots of the log returns against weight showed that a linear relationship existed for all the age-Julian day combinations. This relationship was strongest for the jacks. Correlations of log jack returns versus weight for each Julian day were as high as .89. These plots and the estimates of the parameters from the models showed that the return of jacks is an increasing function of weight at release whereas the return of adults is a decreasing function.

Moreover, there is some tendency for the slopes to approach zero at the later Julian days. This implies that greater returns of jacks may be obtained by early release of larger juveniles, exactly as Bilton et al. (1982) found. In the case of adults, the maximum predicted returns are obtained by late release of small juveniles. This is contrary to the predictions of Bilton et al. (1982) This is probably due to the inclusion of commercial and sport catches in their analysis. The proportion of adult coho caught in the fisheries is about five to ten times greater than the proportion returning to the hatchery. Thus the factors maximizing hatchery returns are masked by those maximizing catches, and the two are different.

The experiment at Rosewall Creek was well designed using classical experimental design concepts. However, it was only carried out for one brood year. As we saw in sections II.2 and II.3, the yearly variations are very large and there may be interactions between the years and the other factors. This means that extrapolation of the results of the model developed here, or the model Bilton et al. (1982), to predict the returns and/or catches of coho salmon in later years, may give poor predictions. Manipulation of hatchery conditions based on these models could even decrease rather than increase the returns and catches for some brood years. Bilton et al. (1982) did some comparisons of their results with those from other hatcheries in Washington State and British Columbia and found a poor correspondence. They concluded that other as yet unrecognized factors may be influencing the results. I agree, and again emphasize the need

for further research to determine what these factors are.

Part III Migration

III.1 Methods for Analyzing Migration Data.

So far this report has been primarily concerned with analyzing perturbation-response relationships and summarizing them with numerical estimates of parameters in statistical models. This approach is of no use when it is impossible to postulate a statistical model which seems reasonable, and when in any case numerical estimates of parameters are not an eloquent way of encapsulating a massive amount of data into a form suitable to the human mind and eye. The role of the statistician in studying the migration of salmon is to find some way of using data not primarily collected for that purpose and summarizing it with some graphical display which shows when the fish went where. Unfortunately, the development of statistics has made little progress in the area of graphics, with a few exceptions, notably the work of Chernoff (1973), and Tukey's (1977) "exploratory data analysis" techniques.

The salmon released from the Robertson Creek hatchery are caught by fishermen in twelve regions along the West Coast of North America

(Figure 2a). Occasionally, the location of capture is not known. Two additional "regions" were created to account for these. The fourteen regions were represented diagrammatically in a 7X2 array (Figure 2b) to facilitate the analysis and avoid drawing many complex maps. The two unknown regions were used in preliminary analyses which are not presented in this report. They are placed in the array in cells which would otherwise have been empty. The array actually removes the visual distraction created by the jagged coastline and allows the eye to concentrate on the important aspects of the data.

The basic characteristics of Pacific Salmon migration are well known. The initial migration from the spawning grounds downstream to the sea, and the return trip from the sea to the spawning grounds, have been well documented in a variety of contexts. The movement of the salmon in the high seas has been less extensively studied, mainly because of the technical difficulties involved. Royce et al. (1968) gave an excellent description of the movements of several species of salmon in the North Eastern Pacific. The movements of chinook salmon are not as well understood as those of other species of salmon.

The data from Robertson Creek CWT groups provide an opportunity to study the migration of chinook salmon. The data include the number of tagged salmon caught each month in each of the twelve regions for five years after the fish were released.

There are three ways in which the fish could be moving. Each individual fish could be moving randomly, as in Brownian motion. This assumption has led to dispersion type models for fish migration, such

as Saila and Shappy (1963), Nicholson and Pope (1977), and others reviewed by Harden Jones (1968). This type of migration is generally not believed to be the case for salmon; rather, it is probably a directed rather than a random migration. Some authors, (eq Royce et al (1968)) believe the migration follows the Alaskan Gyre, a large counterclockwise current circulating in the Gulf of Alaska. Since the juveniles leave the hatchery within a short time interval, and return several years later within a period of several weeks, it is reasonable to assume that they travel in clumps while at sea, rather than dispersing. If this were true, and if the fish were travelling along some axis, then we would naturally expect that fish caught closer to the origin of movement would also be caught sooner than those caught farther away. Since not all fish migrate at exactly the same time there will be some overlap of the distribution of recapture times at the two locations. In order to study where the fish were first, we need some measure of the position of the distribution along the time axis. The three obvious candidates are the mean, mode, and median recapture times.

When the number of recaptures is low there may be more than one mode, so this should not be used. The mean is usually the choice for a location parameter when the distributional form is known to be normal. The median, however, is a more appropriate measure of location when no distributional form is assumed. I plotted the number of recaptures at each month, age and location for the twenty CWT groups and do not feel myself in a position to assume a distributional

form for the recapture times. I therefore chose the median.

The median recapture times for each age and brood year were calculated for the fourteen locations for both troll and net fisheries. A "mosaic" similar to the one used by Hartigan and Kleiner (1984) was chosen to display the medians. In Hartigan and Kleiner's mosaic, the size of each rectangle in the mosaic is proportional to a corresponding cell count in a cross-classified contingency table. In this case, the size of each of the squares in the 7X2 array is proportional to the median recapture time for a certain age and brood year. All CWT groups in each brood year were tallied together, and the medians were calculated. The data were then grouped together across all the brood years and the medians recalculated. This was done to average out the brood year differences. The results are displayed in Figures 4 and 5.

A second mosaic for each of the two fisheries was drawn to display the sample sizes (i.e. number of recaptures) used in calculating the medians (Figures 6 and 7). Unlike Hartigan and Kleiner, the sizes of the boxes are proportional to the logarithms of the counts. The boxes which are less than one-half the dimensions of the corresponding cells represent small sample sizes, in the range of 2-20 recaptures for the troll and 2-15 for the net fisheries. The largest boxes represent approximately 800 and 450, respectively, for the two fisheries. The purpose of these mosaics is to make relative comparisons of the number of recaptures and to help assign subjective weights to the estimates of median recapture times.

The size of the fish at the different locations and ages is also of interest. Figures 8 and 9 were used to display the fish lengths calculated from the troll data. The same 7X2 array was again used. The size of the fish in each cell is directly proportional to the average length of the fish. The smile is a function of the log of the sample size. All smiling fish correspond to a sample size greater than twenty five. Thus smiling fish should be given more weight than frowning fish.

These smiling fish are similar in concept to Chernoff's faces (Chernoff, 1974) where he used faces with eight variable features to represent multivariate data. These fish contain information on average length, sample size and location of recapture in a very compact form. The width of the tail could have been used to represent the spread or dispersion of the measured fish lengths had the raw data been reported. These fish do not graphically display multivariate data; rather the features of the fish are used to display summary data for a large number of observations.

Figure 4. MEDIAN RECAPTURE TIMES BY AGE AND BROOD YEAR (TROLL)

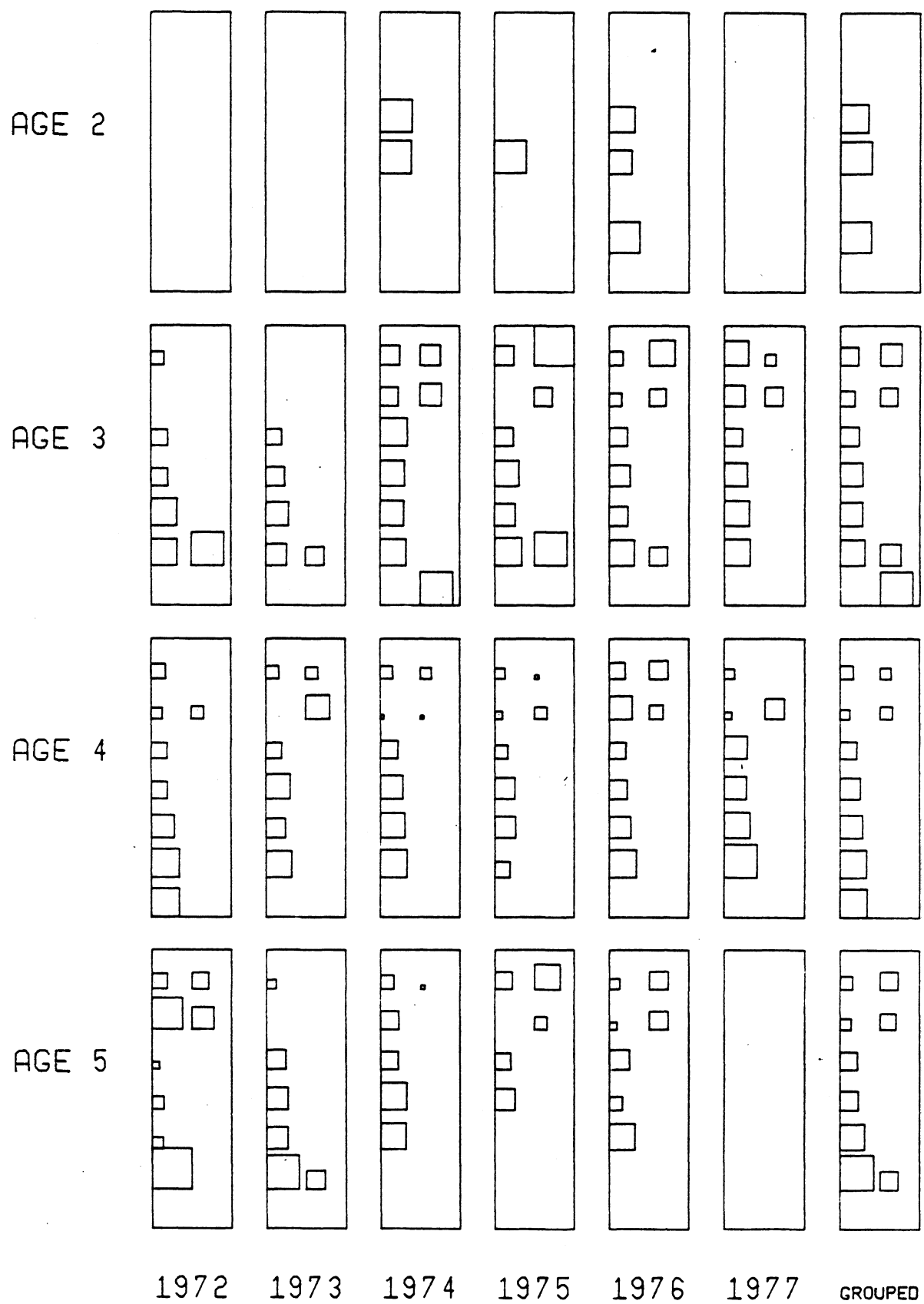


Figure 5. MEDIAN RECAPTURE TIMES BY AGE AND BROOD YEAR (NET)

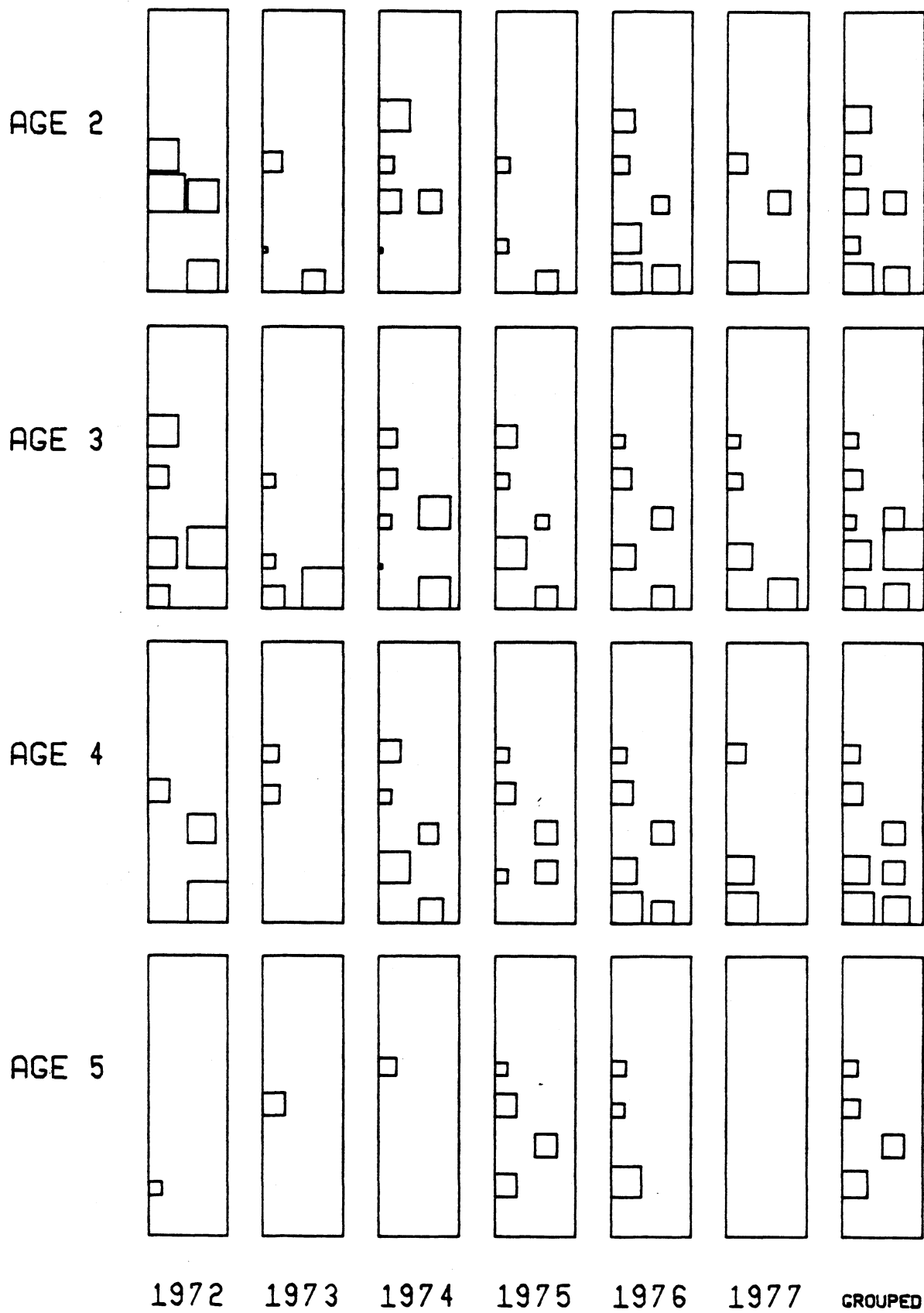


Figure 6. LOG NUMBER OF RECAPTURES BY AGE AND BROOD YEAR (TROLL)

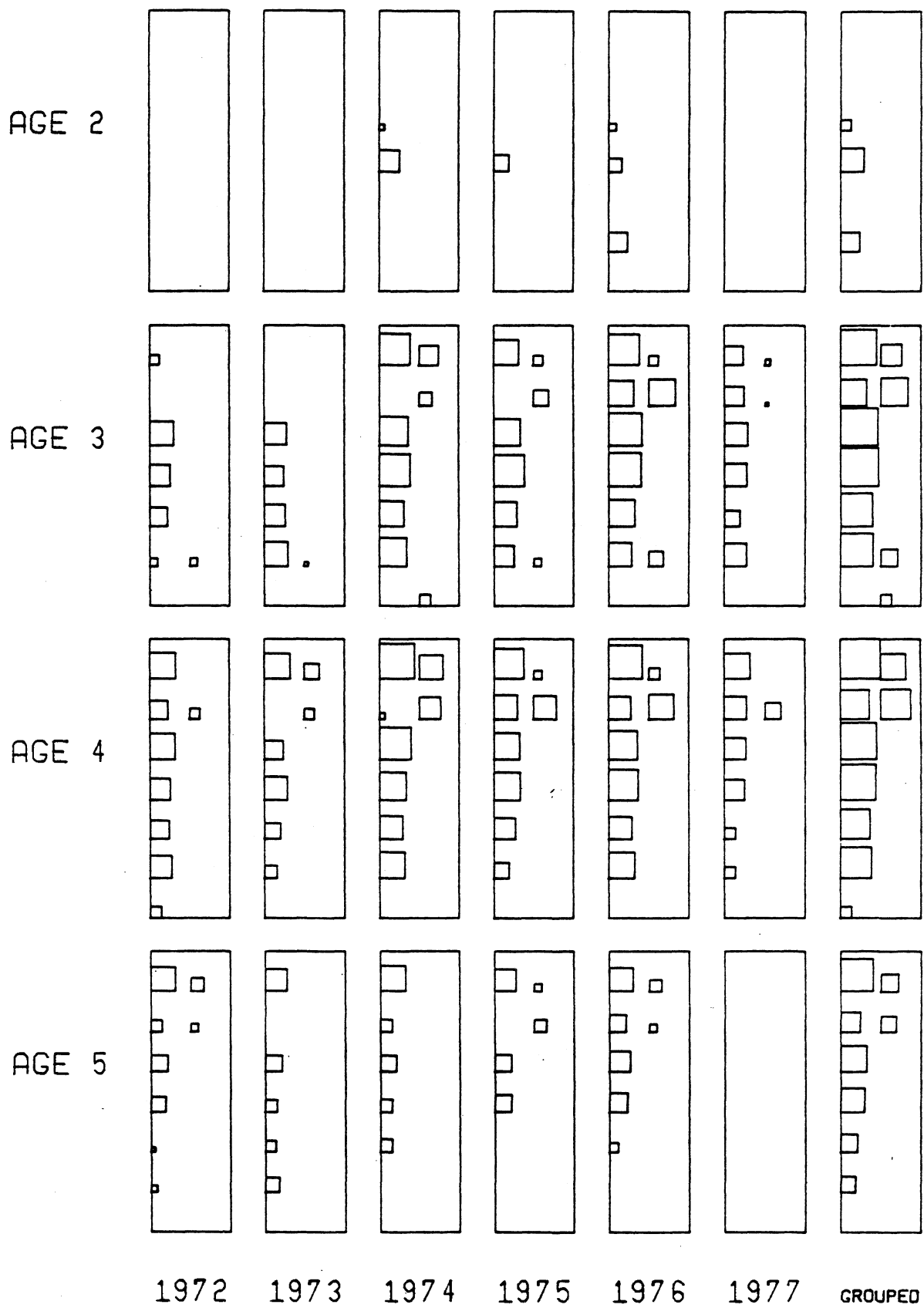
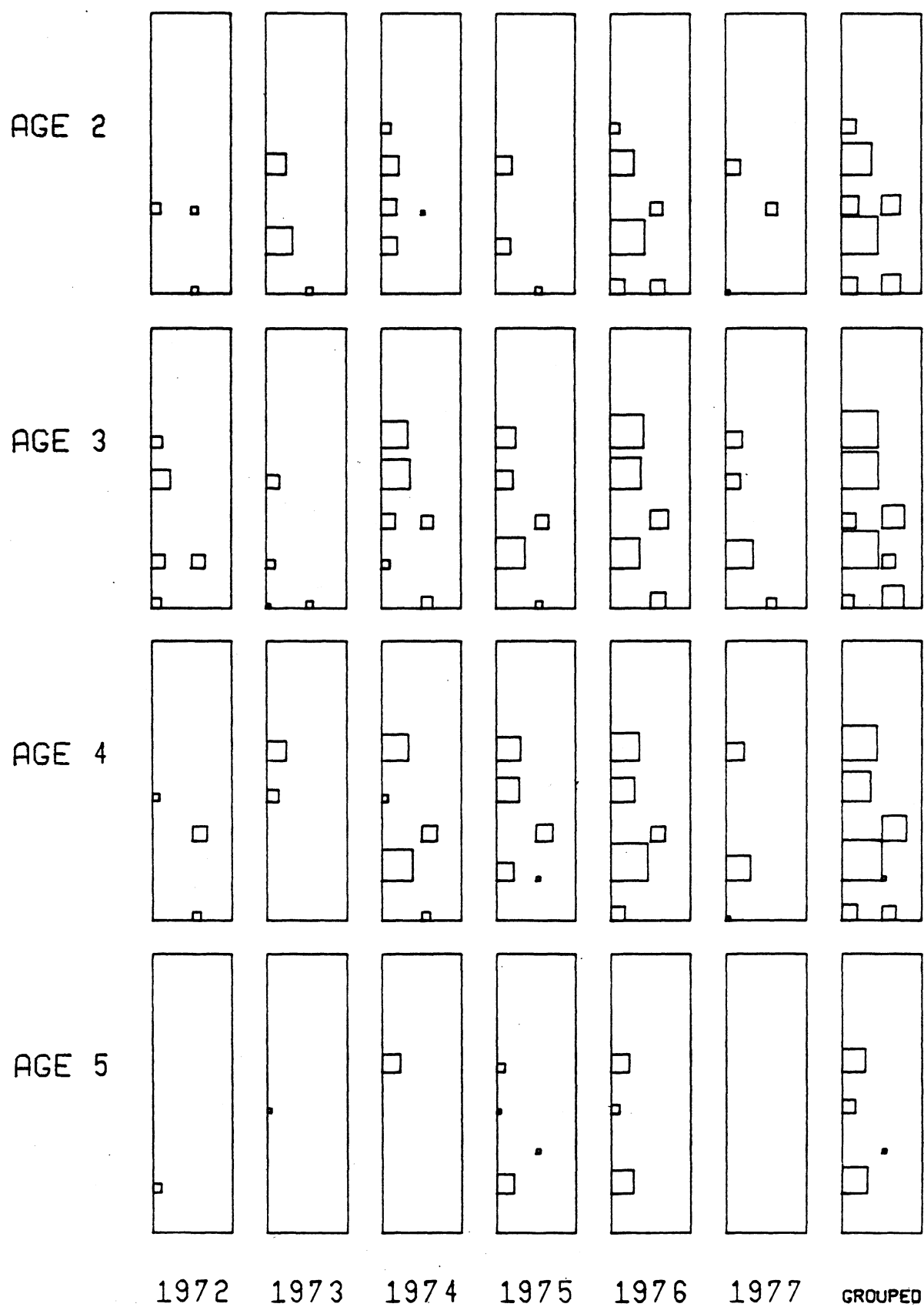
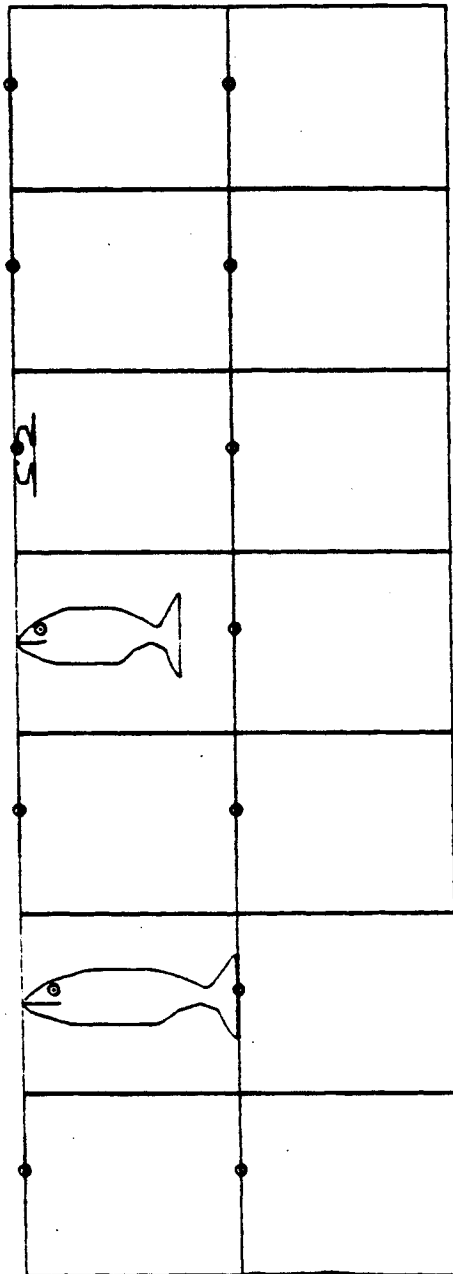


Figure 7. LOG NUMBER OF RECAPTURES BY AGE AND BROOD YEAR (NET)

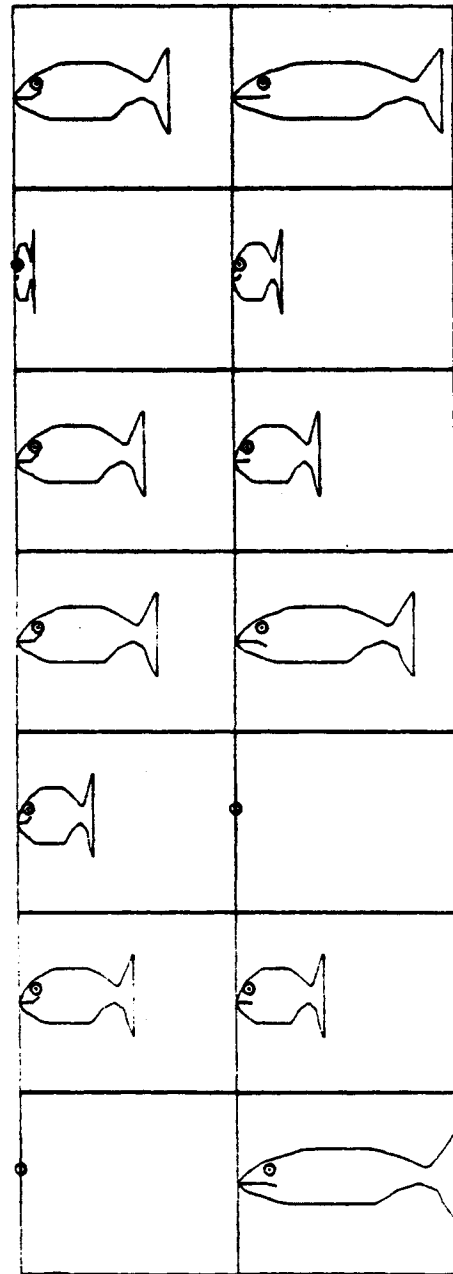


WEIGHTED AVERAGE
LENGTHS, AGE 2



LENGTH RANGE: 48-51

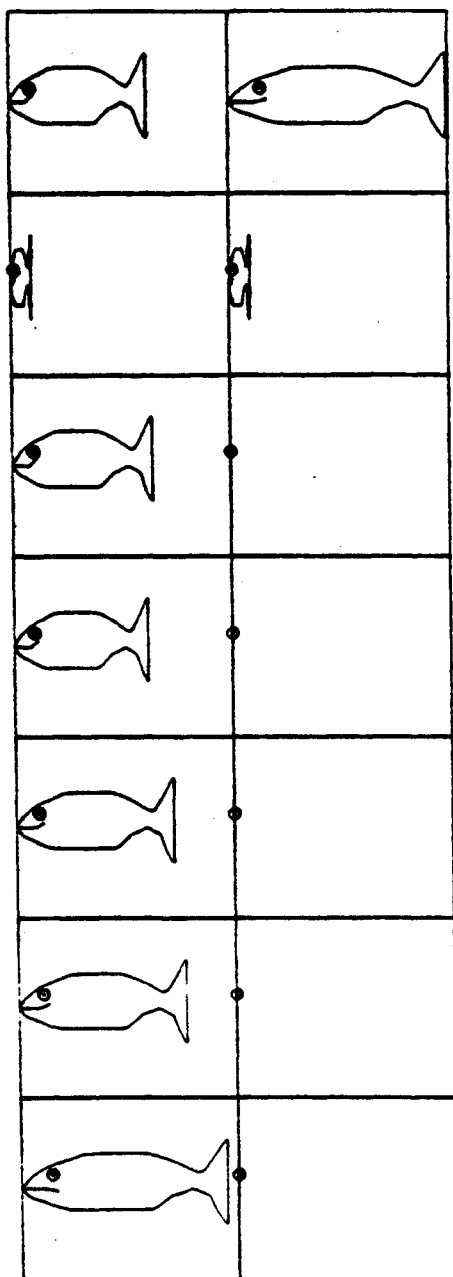
WEIGHTED AVERAGE
LENGTHS, AGE 3



LENGTH RANGE: 62-77

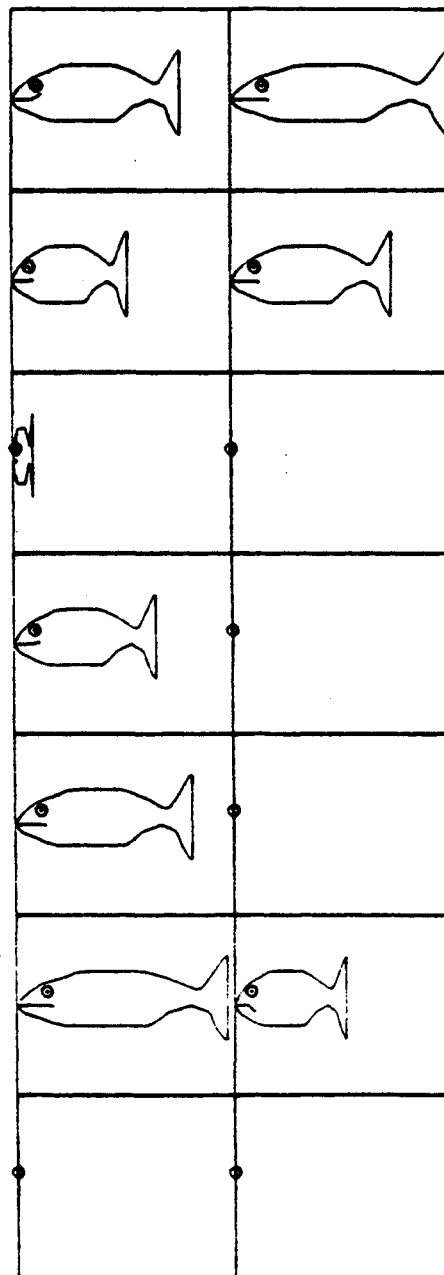
Figure 8.

WEIGHTED AVERAGE
LENGTHS, AGE 4



LENGTH RANGE: 73-87

WEIGHTED AVERAGE
LENGTHS, AGE 5



LENGTH RANGE: 54-91

Figure 9.

III.2 Interpretation of the Mosaics.

When there were recaptures in one of the fourteen regions a box was plotted in the location of the corresponding cell in the 7X2 array. The lower left hand corner of the boxes and cells coincide. The larger the box, the later the median recapture time. Any sequence of boxes from small to large indicates the direction of movement. The smallest box in Figure 4 corresponds to the earliest recapture time, which was on about May 15. This median occurred at age five in Northern B.C. for the 1972 brood year, at age four in South Outside Alaska and at age five in North Inside Alaska for the 1974 brood year, and at age four in North Inside Alaska for the 1975 brood year. The latest median was at age three in North Inside Alaska for the 1975 brood year. This date corresponds to mid-October. However, as we see from Figure 6, the number of recaptures at this time was quite small.

At age two there are no recaptures in Alaska and no apparent trend to later recaptures along any geographical axis. This suggests that there is no strong migration at this age. The maximum of the weighted average lengths of the fish caught in B.C. was 51 cm. which is several centimetres below the size limit for the troll fishery in Alaska. We cannot conclude that at age two there are no Robertson

Creek chinook caught in Alaskan waters.

At age three there are some trends. The earliest recaptures are in Alaska. For the grouped data the earliest recaptures are in South Outside Alaska. The same is true for brood years 1974 and 1976. Although for three of the brood years none were caught in this region. The recapture dates become later when moving from north to south. This is apparent for most brood years but especially for the grouped data. The Georgia Strait recaptures were earlier than most of the B.C. recaptures (except in 1975). However, the sample sizes were fairly small and we must remember that for the troll data, Georgia Strait includes fish caught in Johnstone Strait, which is at the same latitude as North West Vancouver Island.

At age four the same trends are apparent. The recaptures are made earliest in South Outside Alaska (see the grouped data). During each of the brood years the region with the earliest median recaptures varies quite considerably between the four Alaskan regions, but so do the sample sizes. The trend to later recaptures in the southern regions is also clear and the sample sizes are quite high.

At age five the median recapture times vary considerably over the first five brood years (the data are missing for 1977) but this is probably due to the low sample sizes. When the data are grouped the same trends as at ages three and four are apparent.

There is a trend to decreasing recaptures in the southern regions. This must be due to either decreasing fishing effort or decreasing numbers of fish available to be caught. Without the effort

data, it is impossible to deduce which. We would expect some mortality with time as the fish migrate, so to some extent mortality will contribute to the decreasing recaptures.

The net recapture mosaics are not as revealing as the troll mosaics. It is very difficult to notice any trends. At ages three to five for the grouped data there is some evidence of increasing recapture times farther south, but it is not strong. The numbers of recaptures show no trends along any geographical axis but for the grouped data show some consistency from age to age. This could be a reflection of the fact that the net fishery is less size selective and does not depend on the fish's willingness to attack a lure. The fish drawings in Figures 8 and 9 demonstrate some of the same trends noticed in the mosaics. The smallest fish are caught in either South Outside Alaska, Northern B.C. or South Inside Alaska. These are the same regions in which the earliest recaptures occur. Except at age three, the fish become progressively larger towards the southern regions. At all ages (except two when there are no data) the fish caught in Alaska, especially North Inside, are large.

Discussion

Royce et al. (1968) and other authors, such as French and McAlister (1970), think that salmon spend much of their adult lives following the currents in the Alaskan Gyre. Their conclusions are primarily based on data from species other than chinook. The data were collected primarily by doing mark-recapture experiments in which marked adult fish were released on the open ocean away from the coastal fishing regions from which this data set comes.

The Alaskan Gyre reaches the west coast of North America at approximately the southern tip of the Alaskan Panhandle, which corresponds to the South Outside Alaska catch region. From here it splits north and south. If the fish were riding the currents, this would explain why South Outside Alaska was consistently the region with the earliest recaptures times. When the current splits, the fish go in both directions. Those going south migrate to the hatchery, those going north could return to the Alaskan gyre and repeat a cycle, which probably takes one year (Royce et al. 1968). The two and three year olds that return to the hatchery are predominantly male. The four year olds are mixed and the five year olds predominantly female. Sex must therefore govern the direction the fish go in when the gyre reaches the coast, males going south at ages 2-4, females at ages 4-6.

The fish caught in North Inside Alaska are usually as large or larger than those caught in the southern regions. This is so despite the generally earlier recapture times. At age three this might be accounted for by a length difference between males and females. At age five, when the fish in the south are predominantly female, the difference in length is much smaller than at ages three and four, supporting the idea that the two sexes contribute to the observed differences in length. However, if both groups of fish, those going south and those going north, were growing at the same rate, the Alaskan fish should be smaller since they were caught earlier. Since they are larger, it seems reasonable to conclude that the southward moving fish are growing, but at a slower rate. This is probably due to less food consumption, which is generally thought to occur during the home migration. The method of capture (for the troll fishery) involves the fish attacking a lure. If they were eating less, less would be caught: this is reflected in the decreasing troll catches seen in Figure 6. This is not evident in the net catches, which do not depend on the fish's hunger.

It is impossible to make any conclusions on where the fish have been and where they are going at age two. It is not possible to say whether they have already been once around the gyre, whether they are migrating north or south, or whether they are residing along the coast and growing. It will not be possible to obtain hard data on this age group unless the sampling method is changed. Hartt (1980) reports that juvenile salmon from the Columbia River (in Washington) reached

the north eastern Gulf of Alaska (which is the same as the North Outside catch region) by the end of their first summer at sea. If they followed the same pattern as the other ages, these juveniles could circle the Alaskan gyre and some of them then migrate southward as jacks during their second summer. The southward migrating jacks could account for the catches and hatchery returns at age two. Major et al. (1978) believe that chinook are confined to the coastal region during their first year at sea. If this were true, the age two fish could not have been around the gyre, and would be caught close to where they had spent their first year.

If we do not assume that all of the fish are migrating along some axis, then the mosaics may be given a different interpretation. Neave (1964) believes the journeys of the salmon are not closely controlled by currents. He suggests that migrating salmon do not necessarily follow a coastline, but could go directly from the high seas to their home stream. It may be that the recapture times at the southern latitudes are due to fish arriving later via a direct route. This idea would be consistent with Saila and Shappy's (1963) model, or with Harden Jones' (1968) model of linear radial scattering. Both of these models require the assumption that salmon can swim in straight lines over large distances. If these models were good representations of reality, the salmon would have to be moving fairly close together, or would have to aggregate as Harden Jones suggests, before the radial scattering began, otherwise the progression of later recapture times would not be observed. Instead, they would be continuously arriving

at the hatchery and the other regions. In fact, the timing of the arrival of the salmon at their home stream is in general very precise from year to year, and this has been documented for this particular stock (Schnute and Sibert, 1983). It seems difficult to imagine that the salmon could be moving together or have aggregated and then scatter into smaller groups or even individually to give the progression of recapture times observed in these mosaics. The first model, of following the gyre, seems more plausible to me.

The data analysis presented here is not sufficient to determine the migration patterns of chinook salmon in the high seas prior to reaching the coast. It does however provide some indication that the gyre-following models of oceanic migrations of other species of salmon may be applicable to Robertson Creek chinook, and by extension, to other southern stocks.

IV Conclusions

In section II we attempted to build a model which could tell us which factors affected when, where and how many fish were caught. We concluded that the year to year variations were enormous and far outweighed the effects of any manipulations which could be done in the hatchery. In section III we studied the dynamics of the fish migrations. In my view the exploratory data analysis techniques provided much more insight than the "confirmatory" techniques (Tukey, 1977); however the two do complement each other. The results of the exploratory analysis, using mosaics and drawing fish, help explain the results of the statistical models. The perturbations in the hatchery (unless they are extreme, such as severely underfeeding the fish or keeping them in ponds for more than one winter) cannot affect whether fish are caught in Alaska or B.C. because the fish are behaving differently in the two places. The Alaskan fish are growing and eating. The British Columbian fish are reducing their food intake and returning to the hatchery to procreate, thus disadvantaging the British Columbian fishermen. It seems unlikely that the manipulations in the hatchery can affect the migration routes. Increasing the total number of fish taken by fishing is a different question, but again the

hatchery manipulations do not seem to play a dominant role here. In order to gain a better understanding of the factors affecting the catches of chinook salmon, investigations must be carried out to discover the causes of the year-to-year variations.

V Recommendations for Experimental Design

1. Hatchery Experiments

All experiments should have a factorial design which as far as possible is orthogonal, as attempted by Bilton et al. (1982) at Rosewall Creek. The experiments should be carried out with the same design over several brood years and at different hatcheries and measurements of the environmental factors which could be causing variations in catches and returns, such as water levels, temperature, and others believed to have an effect, should be made and recorded each year in an attempt to explain brood year differences. As far as is possible all cells in the design should have data. The number of fish that should be tagged depends on the forecasted sampling effort.

2. Sampling

Two approaches can be taken to increase the number of tagged fish in the samples: increase tagging or increase sampling. Which one to use depends on the relative costs of each to a certain extent. However there are statistical considerations as well. Increasing the number of marks or increasing the sampling effort should both produce the same increase in the number of marked fish recaptured, except that by releasing more marked fish, only those stocks with the extra marks

will show more recaptures if the sampling is not changed. There is an advantage to increasing the sampling effort. The variance of the expanded recaptures is proportional to the square of the expansion factor. When sampling is increased the sampling factor is decreased, the variance decreases even more, not just for Robertson Creek fish but for all stocks. All things being equal, I would clearly recommend increasing the sampling effort.

3. Fishing Effort

As we have seen, the absence of data on fishing effort makes it impossible to draw conclusions on several important matters. These include primarily the causes of yearly variations in catches, and geographical differences. If we wish to predict when and where fish are caught, then we must have some idea of how this is influenced by fishing effort. Unfortunately, effort data are difficult to collect for political as well as for logistic reasons.

4. Data Reporting

The following data, had they been available, would have facilitated the data analysis done in this report:

- expansion factors at each region and month (which should be the same for all CWT codes)

- lengths of all recaptured fish

- lengths and time of arrival of all fish returning to the hatchery.

- lengths greater than 99 cm should be reported. At present the computer program which prints out the recaptures reports fish of 100

cm. or greater as 99 cm. This biases the results of any analyses using lengths.

Appendix A

Weighting the Data

In order to see how the data should be weighted to account for the expansion factors we must enter into a brief discussion on model fitting in exponential families of distributions. The following discussion is adapted from McCullagh and Nelder (1983). Their approach to generalized linear models was used for its flexibility and because its consistency with the computer package GLIM (Generalized Linear Models) produced by the Numerical Algorithms Group (Baker and Nelder, 1978).

A density function of the form

$$f_Y(y; \theta, \phi) = \exp((y\theta - b(\theta)) / (a(\phi)) + c(y, \phi)) \quad (1)$$

is said to be a member of an exponential family of distributions for ϕ

known and a, b and c some functions. The parameter θ is called the dispersion parameter. In the case of the Poisson distribution

$$f_Y(v; \mu) = e^{-\mu} \mu^v / v! \quad (2)$$

which can be re-expressed in the form

$$f_Y(v; \theta, \phi) = \exp((v\theta - e^\theta) - \ln v!) \quad (3)$$

where $\theta = \ln \mu$, $a(\phi) = 1$, $b(\theta) = e^\theta$, $c(y, \phi) = -\ln y!$ and $\phi = 1$

The general form of the log likelihood is then

$$l(\theta; y) = \sum ((v\theta - b(\theta)) / a(\phi) + c(y, \phi)) \quad (4)$$

The "deviance" is defined as

$$\text{deviance} = -2(l(\theta; y) - l(y; y)) \quad (5)$$

where $l(y; y)$ is the maximum likelihood achievable for an exact fit in which fitted values equal the data. The deviance is a measure of the

discrepancy of a fit. For the Poisson distribution, the deviance is

$$\text{deviance} = 2 \left(\sum (y \ln(y/\mu) - (y - \mu)) \right) \quad (6)$$

(with summation over all cells)

which is the statistic labelled G^2 by Bishop et al. (1975) when $a(\phi)=1$.

The variance of the data after the fit can be calculated from the second derivatives of the log likelihood.

$$\text{Var}(Y) = b''(\theta) a(\phi)$$

In the case of the Poisson distribution,

$$\text{Var}(Y) = \mu a(\phi) \quad (7)$$

where μ is the mean.

In this data set the multiplication of the sampled recaptures by an expansion factor means that if we set the function $a = 1$ the variance will be underestimated. To see this, let

Y' = number of sampled recaptures

Y = expanded number of recaptures

X = number of sampled and unsampled recaptures.

Then $Y' = f.Y$ where f is the sampling fraction.

Assume $Y' \sim \text{Poisson}(\mu')$

$X \sim \text{Poisson}(\mu)$

$\mu' = f\mu$

Then $E(Y') = \mu'$, $\text{Var}(Y') = \mu'$

$E(X) = \mu$, $\text{Var}(X) = \mu$ (8)

If we expand the sampled recaptures by the sampling fraction or the inverse of the expansion factor then

$$E(Y) = 1/f E(Y') = (1/f) \mu' = \mu \quad (9)$$

$$\text{but } \text{Var}(Y) = \text{Var}(Y'/f) = 1/(f^2) \mu' = (1/f) \mu \quad (10)$$

By assuming Poisson errors, the implicit assumption is also being made that

$$E(Y) = \text{Var}(Y) = \mu \quad (11)$$

Comparing this to (10) we can see that we are underestimating the variance of the expanded recaptures. To correct this we can set

$$a(\phi) = \phi/f \quad (12)$$

where $\phi = 1$ during model fitting but can be estimated by

$$\hat{\phi} = G^2/(N-p) \quad (13)$$

when a satisfactory fit is found. Using (12), we have

$$\text{Var}(Y) = \mu a(\phi) = 1/f \mu \quad (14)$$

which agrees with (10). The deviance will be decreased by the use of (12) as we can see by comparing (4) and (5).

The assumption of Poisson errors is justified on the grounds that we are dealing with counts where the probability of being captured (i.e. counted) is small. The maximum percent recapture (of those released) of the twenty CWT groups, where recaptures were summed over all regions and years, was only 5%. However, since $\text{Var}(Y) = E(Y)$ we will not assume a Poisson distribution but only that

$$\text{Var}(Y) = a(\phi)E(Y) \quad (15)$$

Thus we assume that the errors are distributed as a scaled Poisson.

In most log-linear models there are no replicates; in this case there are cells with several counts in them. This adds to the over-dispersion because the only way to account for differences in replicate counts would be to include a replicate "effect" in the model, which would make the model of no use for prediction.

Effects of Expansion Factors on the Poisson Distribution

A small simulation study was done to examine the effect of the expansion factors on the Poisson distribution. Figures 10-15 show the effect of scaling the recaptures in the sample with the expansion factors. In each figure there are four plots, which were generated using IMSL routines on the Cyber 730 as follows.

The top plot of each Figure is a frequency histogram of 1,000 Poisson deviates with the mean in brackets (for example in Figure 10 the mean is 5.0). This plot is meant to represent the distribution of marked fish in the fishery, that is, in all the boats.

The second plot represents the distribution of the expanded recaptures. To calculate this, the numbers of recaptures (which are on the x axis, not to be confused with the frequency with which those numbers of recaptures were obtained, which is the y axis) in the first plot were divided by the expansion factor and rounded to the nearest

integer, giving the third plot. The new numbers of recaptures were then multiplied by the expansion factor again, to give the second plot. Thus the second plot represents the frequency distribution of the expanded recaptures, and the third plot represents the frequency distribution of recaptures in the sample.

The fourth plot is a frequency histogramme of 1,000 Poisson deviates with the same mean as the recaptures in the sample. It is meant for comparison with the third plot.

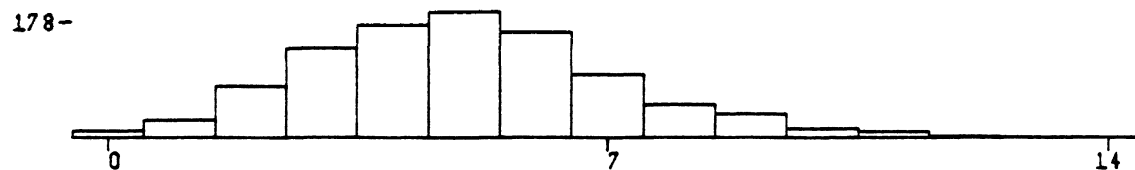
Three expansion factors were used (2.0, 3.5, 8.0) and two means (5 and 250) for recaptures in the fishery.

In all cases the effect of the expansion factor is to leave gaps in the frequency distribution. These gaps are due to the rounding off. The rounding is necessary since only whole fish can be caught!

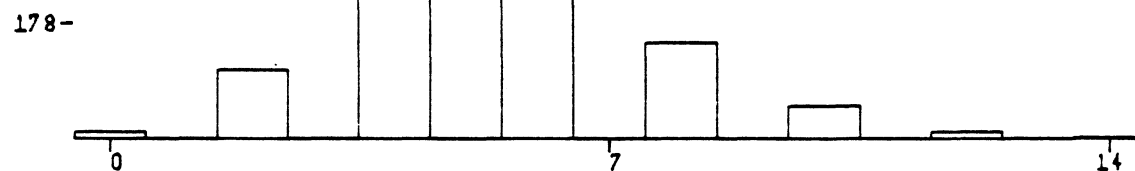
The differences between the four distributions are most noticeable when the expansion factor is high (Figures 14 and 15). When the mean is low and the expansion factor high, the resulting distribution of expanded recaptures represents the true distribution very poorly (Figure 14, plot 1).

When the mean number of marked fish in the fishery is high, the expansion factor makes much less difference. The average expansion factor in the Robertson Creek data set is approximately 3.5, so Figures 12 and 13 show the amount of distortion that could be present in this data set.

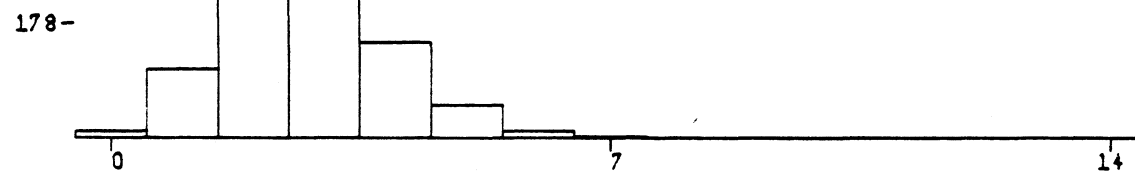
FREQUENCY DISTRIBUTION OF TOTAL RECAPTURES ASSUMING POISSON (5.00)
 357-



FREQUENCY DISTRIBUTION OF EXPANDED RECAPTURES (EXP. FAC. = 2.00)
 357-



FREQUENCY DISTRIBUTION OF RECAPTURES IN SAMPLE (MEAN = 2.50)
 357-



FREQUENCY DISTRIBUTION OF POISSON DEVIATES (MEAN = 2.50)
 357-

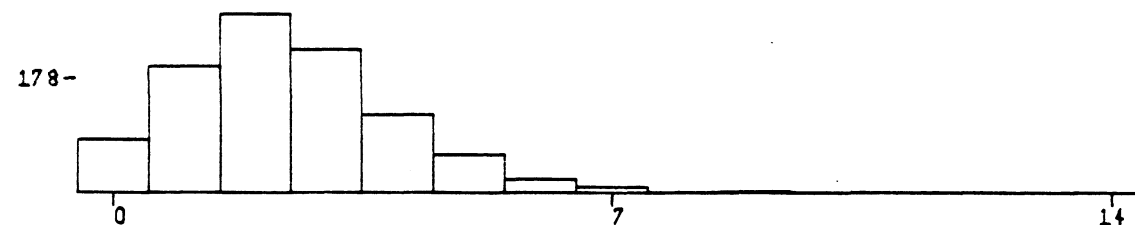


Figure 10.

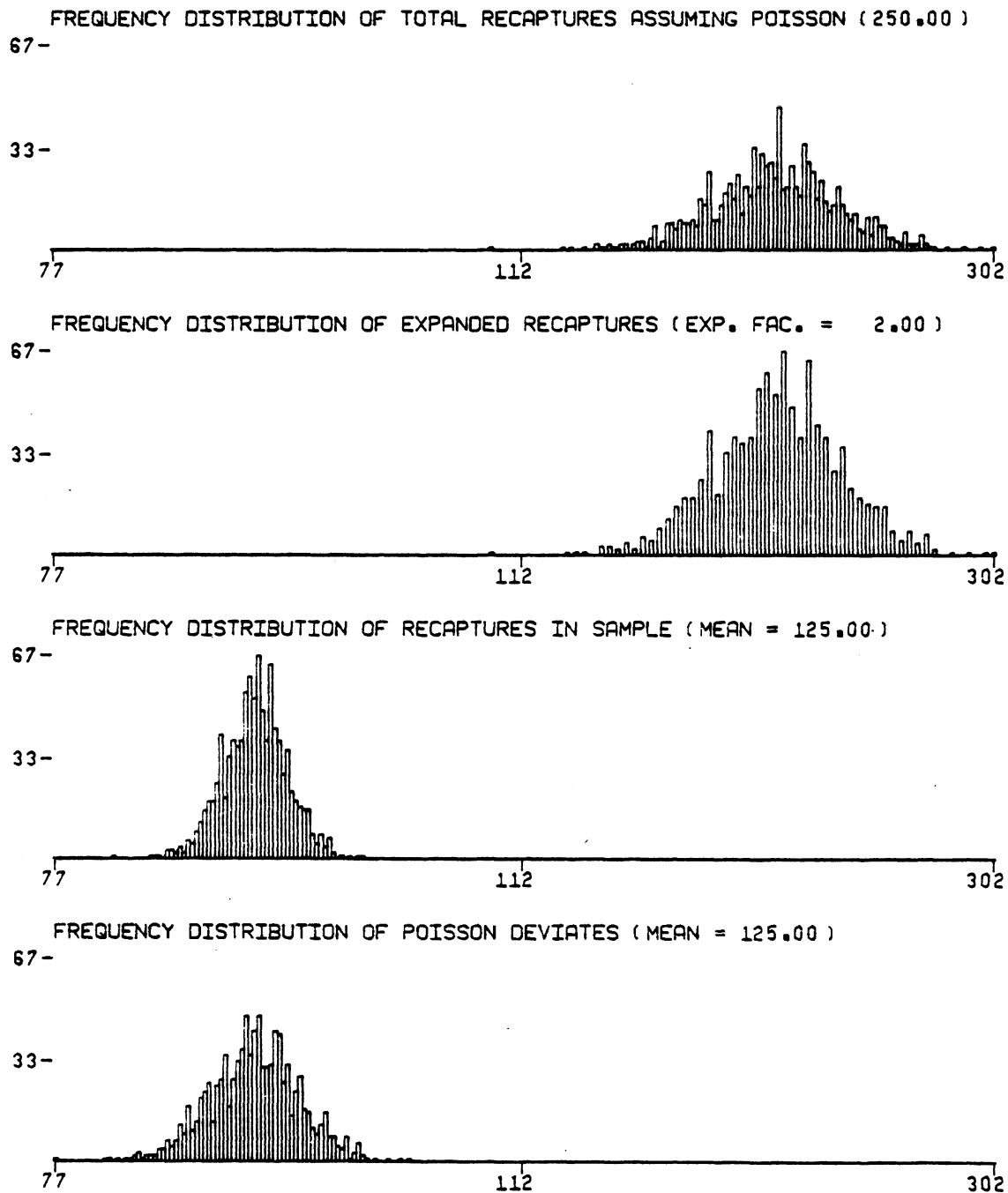
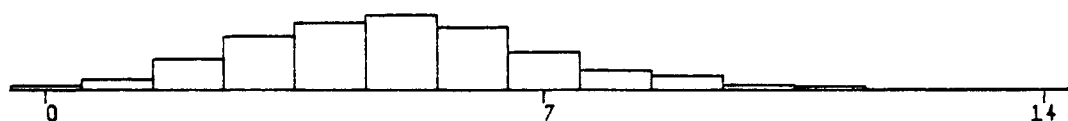


Figure 11.

586- FREQUENCY DISTRIBUTION OF TOTAL RECAPTURES ASSUMING POISSON (5.00)

293-



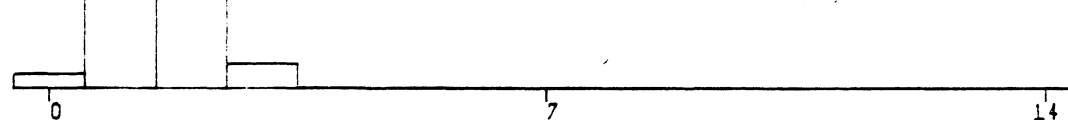
586- FREQUENCY DISTRIBUTION OF EXPANDED RECAPTURES (EXP. FAC. = 3.50)

293-



586- FREQUENCY DISTRIBUTION OF RECAPTURES IN SAMPLE (MEAN = 1.43)

293-



586- FREQUENCY DISTRIBUTION OF POISSON DEVIATES (MEAN = 1.43)

293-

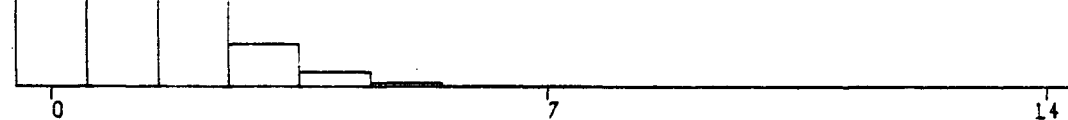
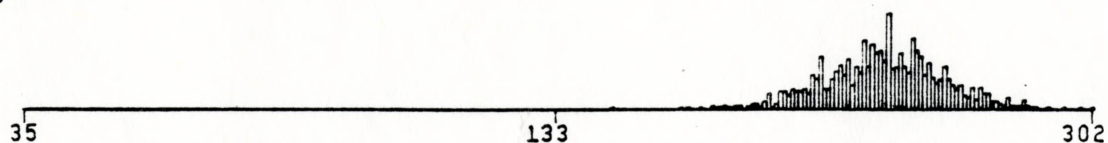


Figure 12.

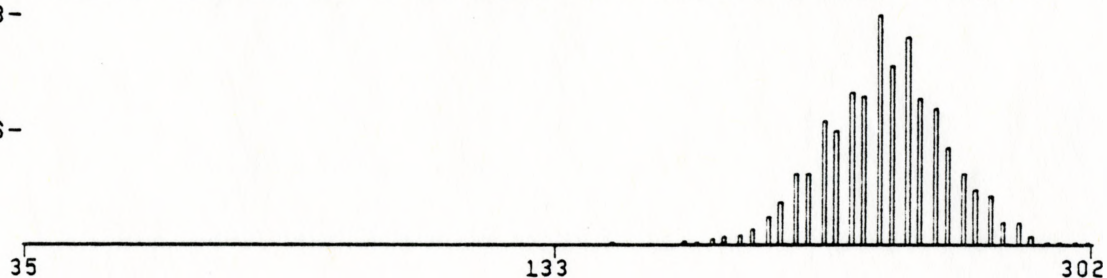
FREQUENCY DISTRIBUTION OF TOTAL RECAPTURES ASSUMING POISSON (250.00)
113-

56-



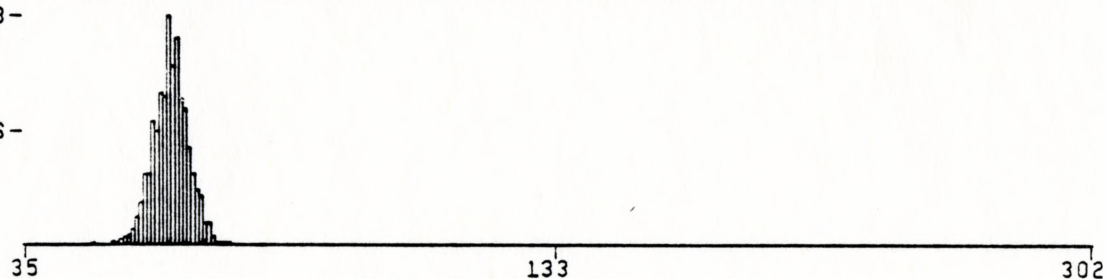
FREQUENCY DISTRIBUTION OF EXPANDED RECAPTURES (EXP. FAC. = 3.50)
113-

56-



FREQUENCY DISTRIBUTION OF RECAPTURES IN SAMPLE (MEAN = 71.43)
113-

56-



FREQUENCY DISTRIBUTION OF POISSON DEVIATES (MEAN = 71.43)
113-

56-

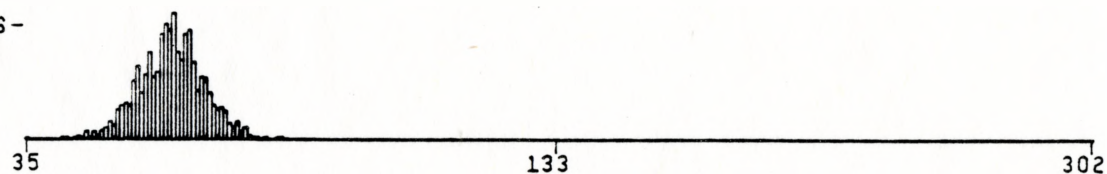
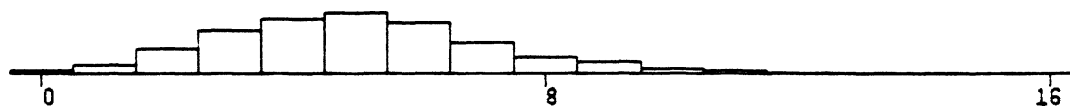


Figure 13.

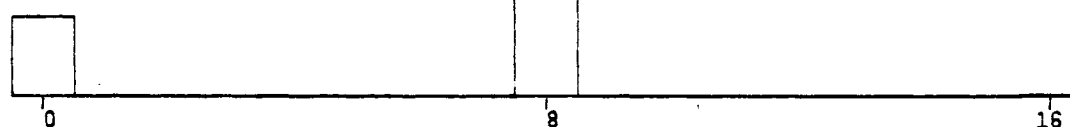
FREQUENCY DISTRIBUTION OF TOTAL RECAPTURES ASSUMING POISSON (5.00)
740-

370-



FREQUENCY DISTRIBUTION OF EXPANDED RECAPTURES (EXP. FAC. = 8.00)
740-

370-



FREQUENCY DISTRIBUTION OF RECAPTURES IN SAMPLE (MEAN = .63)
740-

370-



FREQUENCY DISTRIBUTION OF POISSON DEVIATES (MEAN = .63)
740-

370-

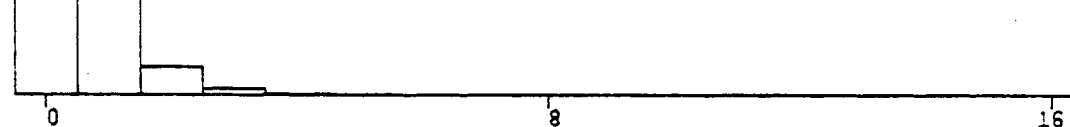
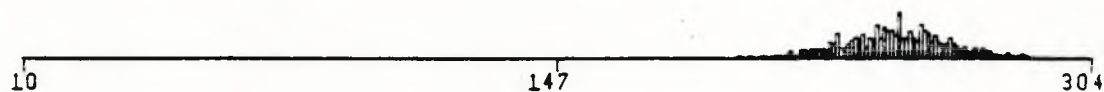


Figure 14.

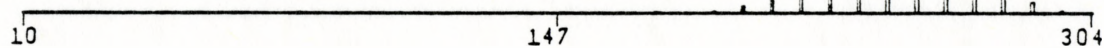
FREQUENCY DISTRIBUTION OF TOTAL RECAPTURES ASSUMING POISSON (250.00)
233-

116-



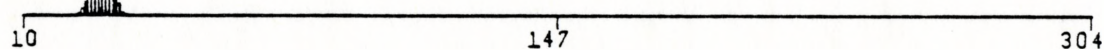
FREQUENCY DISTRIBUTION OF EXPANDED RECAPTURES (EXP. FAC. = 8.00)
233-

116-



FREQUENCY DISTRIBUTION OF RECAPTURES IN SAMPLE (MEAN = 31.25)
233-

116-



FREQUENCY DISTRIBUTION OF POISSON DEVIATES (MEAN = 31.25)
233-

116-

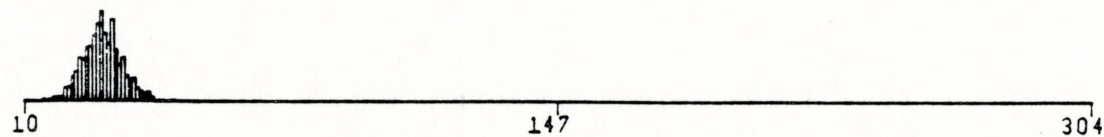


Figure 15.

Appendix B

Description of Coded Wire Tag Groups

(Please refer to table 6)

Explanation

The first column, group number, is the number referred to in the text. The tag code is the binary code written on the wire tags. When fish with missing adipose fins were recaptured or returned to the hatchery, the tags are dissected out from the nose cartilage and this code is read by a technician. The average weight refers to the average weight of a sample of fish taken from the CWT group at the time of release. Julian day of release is counted from January 1 of the brood year and takes leap years into account.

There is some confusion over the treatments the fish were given. The treatments EWOS and OMP refer to two different food types; however some sources say that groups 2 and 3 were production groups, which are the usual, non-experimental, groups. Groups 13-18 were given

treatments early, mid and late release. This is obviously very similar to the Julian day of release. However the mid treatment group in 1976 (#17) was released on the same Julian day as the early treatment group in 1975 (#15) so these treatments must refer to something other than the actual day of release, such as the relative number of days of incubation. A measure of incubation and pond rearing time would be more useful than the day of release, since the 'ideal' day of release probably depends on many factors, such as degree days accumulated, which vary from year to year.

Groups five and twelve were not included in many of the analyses because they were found to be outliers due to their very high and low release weights respectively.

Table 6

Description of the coded wire tag groups (Robertson Creek)

Group Number	Tag Code	Treatment	Release Weight	Number Tags	Total Release	Release Day	Brood Year
1	02/05/01	production	5.7	26881	510739	129	1972
2	02/08/01	ewos	5.5	2627	2627	171	1972
3	02/11/01	omp	5.4	39013	222819	171	1972
4	02/02/03	production	5.9	23480	603717	191	1973
5	02/08/06	supersmolt	41.3	20777	73094	448	1973
6	02/06/02	production	5.0	16137	29352	172	1973
7	02/05/06	production	6.3	27726	520930	196	1973
8	02/04/06	production	6.3	27444	438094	191	1973
9	02/06/06	production	5.2	46194	1003877	162	1974
10	02/09/06	production	5.3	27383	453766	162	1974
11	02/12/06	production	4.5	11674	11674	158	1974
12	02/14/06	production	2.7	28283	228940	162	1974
13	02/04/09	late	5.2	47724	462985	162	1975
14	02/04/08	mid	5.2	50731	487145	155	1975
15	02/13/05	early	5.4	10702	10829	146	1975
16	02/16/29	early	5.3	59227	63054	130	1976
17	02/16/30	mid	5.4	64550	3188072	146	1976
18	02/16/31	late	5.3	69203	442487	159	1976
19	02/22/17	production	4.4	70816	4178531	156	1977
20	02/22/18	production	5.3	66725	3553187	154	1977

Total: 737297 16485919

Appendix C

Rosewall Creek Data Set

The data set consisted of six pond populations, three size groups (small, medium and large) and four release dates, giving a 6X4X3 factorial design. The release dates were Julian days 104, 132, 161 and 189, all in the year 1975. On Julian day 189 there was only one pond population released. If the size groups were considered levels of a factor the design would have to be nested as the as all three size groups increased with the day of release. Since the average weight of each tag group was known, weight was treated as a covariable. Bilton et al. (1982) modelled both weight and Julian day as covariables, but I have treated Julian day as a factor. This gives a 6X4 design with five empty cells and three replicates in all other cells. For a complete description of the data set, the reader is referred to Bilton (1980).

References

- Baker, R.J. and Nelder, J.A. (1978). The GLIM System, Release 3, Generalized Linear Interactive Modelling, Numerical Algorithms Group, Oxford.
- Bilton, H.T. (1980). Returns of adult coho salmon in relation to mean size and time at release of juveniles to the catch and escapement. Can. Tech. Rep. Fish. Aquat. Sci. 941
- Bilton, H.T., Alderdice, D.F. and Schnute, J.S. (1982). Influence of time and size at release of juvenile coho salmon (*Oncorhynchus kisutch*) on returns at maturity. Can. J. Fish. Aquat. Sci. 39:426-447.
- Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). Discrete Multivariate Analysis: Theory and Practice, MIT Press, Cambridge, Massachusetts.
- Brownie, C., Anderson, D.R., Burnham, K.P. and Robson, D.S. (1978). Statistical Inference from Band Recovery Data: A Handbook. U.S. Fish. and Wildl. Serv., Resour. Publ. No 131, 212pp
- Chapman, D.G. (1961). Statistical Problems in Dynamics of Exploited Fisheries Populations. Proc. 4th Berkeley Symp. 1960, 4:153-160.
- Chernoff, H. (1973). The Use of Faces to Represent Points in k-Dimensional Space Graphically. J. Amer. Statist. Assoc.

68:361-368

Cochran, W.G. (1977). Sampling Techniques, third edition. Wiley, New York.

Cormack, R.M.(1981). Loglinear Models for Capture-Recapture Experiments on Open Populations. In The Mathematical Theory of the Dynamics of Biological Populations II (R.W. Hiorns and D. Cooke, eds). pp197-215. Academic Press, London.

Fienberg, S.E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. Biometrika 59, 591-603.

French, R.R. and McAlister, W.B. (1970). Winter Distribution of Salmon in Relation to Currents and Water Masses in the Northeastern Pacific Ocean and Migrations of Sockeye Salmon. Trans. Amer. Fish. Soc. 99 649-663.

Gulland, J.A. (1955). On the estimation of population parameters from marked numbers. Biometrika 42, 269-270.

Harden Jones, F.R. (1968). Fish Migration. Arnold, London.

Hartigan, J.A. and Kleiner, B. (1984). A Mosaic of Television Ratings. Amer. Statis. 38(1), 32-35.

Hartt, A.C. (1980). Juvenile Salmonids in the Oceanic Ecosystem - the Critical First Summer. In Salmonid Ecosystems of the North

Pacific, pp25-57, Oregon State University.

Kalbfleisch, J.D. and Prentice, R.L. (1980). The Statistical Analysis of Failure Time Data. Wiley, New York

Lawless, J.F. (1982). Statistical Models and Methods for Lifetime Data. Wiley, New York

Major, R.L., Ito J., Ito, S. and Godfrey, H. (1978). Distribution and abundance of chinook salmon in offshore waters of the North Pacific Ocean. Int. North. Pac. Fish. Comm. Bull. 38.

McCullagh, P. and Nelder, J.A. (1983). Generalized Linear Models. Chapman and Hall, London.

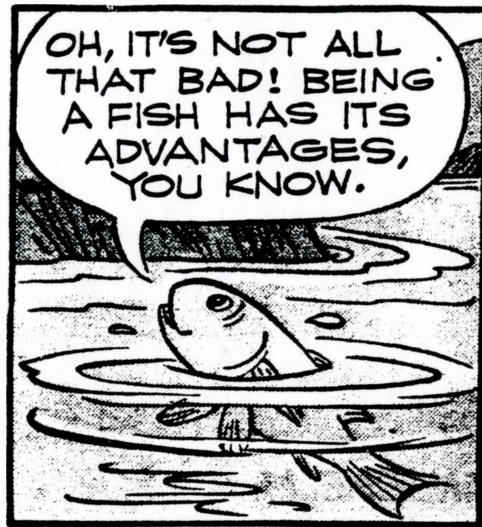
Neave, F. (1964). Ocean Migrations of Pacific Salmon. J. Fish. Res. Bd. Can. 21(5):1227-1244.

Nicholson, M.D. and Pope, J.A. (1977). The Estimation of Mortality from Capture-Recapture Experiments. In Fisheries Mathematics (J.H. Steele, ed.) pp77-85. Academic Press, London.

Pearse, P.H. (1982). Turning the Tide, A New Policy for Canada's Pacific Fisheries. The Commission on Pacific Fisheries Policy, Final Report. Department of Supply and Services, Ottawa, Canada.

Ricker, W.E. (1976). Review of the Rate of Growth and Mortality of Pacific Salmon in Salt Water, and Non Catch Mortality Caused by Fishing. J. Fish. Res. Bd. Can. 33:1483-1524

- Royce, M.P., Smith, L.S. and Hartt, A.C. (1968). Models of Oceanic Migrations of Pacific Salmon and Comments on Guidance Mechanisms. U.S. Fish and Wildl. Serv. Fish. Bull. 66(3):441-462
- Saila, S. B. and Shappv, R.A. (1963). Random Movement and Orientation in Salmon Migration. J. Cons. perm. int. Explor. Mer. 28:153-66
- Schnute, J. and Sibert, J. (1983). The salmon terminal fishery: a practical, comprehensive timing model. Can. J. Fish. Aquat. Sci. 40:835-853
- Seber, G.A.F. (1982). The Estimation of Animal Abundance and Related Parameters, second edition. Macmillan, New York.
- Tukey, J.W. (1977). Exploratory Data Analysis. Addison-Wesley, Reading, Massachusetts.
- Urquhart, N.S. and Weeks, D.L. (1978). Linear models in messy data: some problems and alternatives. Biometrics 34(4):696-705



$$f_Y(y; \mu) = e^{-\mu} \mu^y / y!$$