APPENDIX A

Required Level of Effort of Sampling at the Delta Fish Protective Facility

prepared by

David T. Hughes

Department of Mathematics

University of the Pacific

I. Introduction

It is desired to determine the level of effort necessary to estimate the total number of fish salvaged at the Delta Fish Protective within certain limits (± 50% for counts over 10,000 and ± 100% for counts under 10,000) at the 80% confidence level. Secondary considerations involve the subsampling for species composition and length.

II. Considerations Based on Available Data

The recommendations contained in this report are based on data contained in

1) "Evaluation Testing Program Report for Delta Fish Protective Facility, State Water Facilities, California Aquaduct, North San Joaquin Division," State of California, The Resources Agency, Department of Water Resources, Department of Fish and Game, Memorandum Report, 1973
2) Monthly reports of the fish salvage program (January, 1973 through August, 1976)
3) Daily actual and projected totals of fish salvaged at the Delta Fish Protective for May, July, 1974.

The data contained in these reports indicate that there is considerable seasonal variation over the calandar year. The trend of this variation appears roughly the same from year to year with a period of relative abundance from May through August, a period of relative scarcity from February through April, and a period of a "moderate " number of fish salvaged from September through January. In addition to this seasonal variation, there is considerable variation in the number of fish salvaged during various time periods throughout the day. The data contained in the reports of daily actual and projected totals of fish salvaged for May and July, 1974 indicate a general division of the day into two periods during which there is approximately equal frequency of fish entering the facility:

Period 1 : 2000 hours to 0800 hours (high frequency)

Period 2 : 0800 hours to 2000 hours (low frequency)

(This division is somewhat arbitrary but reasonable based on the data available. However, there is some arguement for including the 2000 to 2200 hour period in with Period 2. I have chosen to place this block in with Period 1 to obtain two periods of equal duration.) The data of daily actual and projected totals indicate that the frequency of fish entering the facility during Period 1 is from four times as great as the frequency during Period 2 (May, 1974) to fifteen times as great as the frequency during Period 2 (July, 1974). This ratio is probably quite different for other months, but nonetheless indicates that these two time periods be dealt with separately.

The nature of the variatin in number of fish salvaged per minute appears to be negative binomial (i.e. $\sigma^2 \div \mu^2$). This is somewhat difficult to determine from the available data since the duration of the sampling periods varies. At any rate, this should serve as a conservative estimate.

III. Recommendations

A. Estimation of Total Salvage and Level of Effort Required

Presently, fish are diverted into a separate holding tank at two hour intervals for the purpose of sampling. The duration of these sampling periods varies according to the number of fish diverted. Statistically, it is more convenient if the duration of the sampling times is constant, at least within a daily time period, although this may not be feasible in the field. With this in mind and the day divided into two time periods as described in section II, let

P_1 = daily time period 1 (2000 to 0800 hours)
P_2 = daily time period 2 (0800 to 2000 hours)
n_i = number of sampling times during P_i (i=1,2)
d_i = duration (in minutes) of each sampling time during P_i (i=1,2)
μ_i = mean # of arrivals/minute during P_i (i=1,2)
σ_i^2 = variance of the # of arrivals/minute during P_i (i=1,2)
y_{ij} = actual # of fish observed during the j^{th} sampling time during P_i (i=1,2)
$T_i = \sum_j y_{ij}$ = total # of fish observed during P_i (i=1,2)
$\hat{T}_i$ = estimate of the total # of fish salvaged during P_i (i=1,2)
$\widehat{GT} = \hat{T}_1 + \hat{T}_2$ = estimate of the total # of fish salvaged during a day

With the above definitions, estimates of the number of fish salvaged and their respective variances are (assuming $\sigma_i^2 = \mu_i^2$)

$$\hat{T}_i = \frac{720 T_i}{n_i d_i} \qquad \mathrm{Var}(\hat{T}_i) = \frac{(720)^2 \mu_i^2}{n_i d_i} \approx \frac{(\hat{T}_i)^2}{n_i d_i}$$

$$\widehat{GT} = \hat{T}_1 + \hat{T}_2 \qquad \mathrm{Var}(\widehat{GT}) = \mathrm{Var}(\hat{T}_1) + \mathrm{Var}(\hat{T}_2)$$

$$\approx \frac{\hat{T}_1^2}{n_1 d_1} + \frac{\hat{T}_2^2}{n_2 d_2}$$

Confidence intervals (80% level) for the total number of fish salvaged during P_i and during the entire day would be:

$$\hat{T}_i \pm 1.28\sqrt{\frac{\hat{T}_i^2}{n_i d_i}} \quad \text{or} \quad \hat{T}_i \pm \frac{1.28}{\sqrt{n_i d_i}}\hat{T}_i$$

and

$$\widehat{GT} \pm 1.28\sqrt{\frac{\hat{T}_1^2}{n_1 d_1} + \frac{\hat{T}_2^2}{n_2 d_2}}$$

Thus, to achieve the desired level of accuracy,we require

$$\frac{1.28}{\sqrt{n_i d_i}} = .5$$

or

$$n_i d_i \approx 6.5$$

The May, July data for daily actual and projected totals indicate an average duration of

$$d_1 = .25 \text{ minutes}$$

$$d_2 = 3 \text{ minutes.}$$

Assuming this to be typical of months of relative abundance, this would require

n_1 = 26 sampling times during P_1

n_2 = 3 sampling times during P_2.

During months of less abundance, the d_i's could be lengthened with correspondingly fewer sampling times. The following table gives possible values of d_1 and d_2 and the corresponding number of sampling times. I have arbitrarily assumed a minimum number of n_i = 3 sampling times during a time period.

Table 1

Duration and number of sampling times during period P_i $(i=1,2)$ $(n_i$ 3)

d_i (minutes)	.25	.5	1	2	3	3
n_i	26	13	7	4	3	3

Example: This example is based on the data from the July, 1974 report of daily actual and projected totals of fish salvaged at the Delta Fish Protective Facility. Assuming mean numbers of fish salvaged of

λ_1 = 307/minute during P_1

λ_2 = 20/minute during P_2

with durations of

d_1 = .25 minutes for n_1 = 26 sampling times

d_2 = 3 minutes for n_2 = 3 sampling times,

we would expect to observe a total of

$T_1 \approx n_1 d_1 \lambda_1 = (26)(.25)(307) = 1996$ fish during P_1

$T_2 \approx n_2 d_2 \lambda_2 = (3)(3)(20) = 180$ fish during P_2

giving estimates of

$$\hat{T}_1 = \frac{720}{n_1 d_1} T_1 = 221{,}096$$

$$\hat{T}_2 = \frac{720}{n_2 d_2} T_2 = 14{,}400$$

and

$$\hat{GT} = \hat{T}_1 + \hat{T}_2 = 235{,}496$$

Corresponding 80% confidence intervals would be

a) During P_1

$$\hat{T}_1 \pm \frac{1.28}{\sqrt{n_1 d_1}} \hat{T}_1$$

221,096 ± 111,003

b) During P_2

$$\hat{T}_2 \pm \frac{1.28}{\sqrt{n_2 d_2}} \hat{T}_2$$

14,400 ± 6,144

c) During the entire day

$$\widehat{GT} \pm 1.28 \sqrt{\frac{\hat{T}_1^2}{n_1 d_1} + \frac{\hat{T}_2^2}{n_2 d_2}}$$

235,496 ± 111,173

B. Subsampling for Species Composition and Length

No specific recommendations are given here, rather, some general considerations are laid out. Based on data obtaind from the 1973 memorandum on the evaluation testing program, variation in lengths of the four species of primary interest (striped bass, white catfish, king salmon and American shad) is relatively small ($\sigma \approx 20$ mm for SB, WCF and KS and $\sigma \approx 10$mm for AS). Accuracy of length estimates is determined by

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}} \approx \frac{20\text{mm}}{\sqrt{n}} \quad \text{for SB,WCF and KS}$$

$$\approx \frac{10\text{mm}}{\sqrt{n}} \quad \text{for AS}$$

where n represents the number of fish within a species on which the length estimate is based. This will generally be different for different species.

Accuracy of the estimates of species composition depend upon the percent, P, of the species present with a standard deviation of

$$\sigma_p = \sqrt{\frac{P(100-P)}{N}}$$

where N is the sample size on which the estimate is based. This is maximum when P=50% so that

$$\sigma_p \leq \frac{50\%}{\sqrt{N}}$$

The number of fish within a species will be small relative to the total number of fish sampled to estimate length and species composition. Thus the subsampling procedure should be such that the number of fish within a species is sufficiently large to insure the desired accuracy of length estimates. Sample sizes within a species of n = 100 would yield length estimates accurate to approximatly (with 95% confidence)

$$2\sigma_{\bar{y}} = 2\frac{\sigma}{\sqrt{n}}$$

$$2\,\frac{20\text{mm}}{\sqrt{100}} = 4\text{mm} \quad \text{for SB,WCF and KS}$$

$$2\,\frac{10\text{mm}}{\sqrt{100}} = 2\text{mm} \quad \text{for AS}$$

Since the percentage of a species present, as well as the number of fish passing through the facility, varies considerably, the frequency of subsampling could be adjusted accordingly. Less frequent subsampling could occur during months of relative abundance with more frequent subsampling during months of relative scarcity.