A Statistical Model for the Survival of Chinook Salmon Smolts Outmigrating Through the Lower Sacramento-San Joaquin System

Ken Newman and John Rice
Division of Statistics, University of Idaho
Department of Statistics, University of California, Berkeley

November 19, 1997

U.S. Fish and Wildlife Service provided financial support for Newman and California Urban Water Agencies provided financial support for Rice.

Abstract

A statistical model was constructed for the survival of juvenile chinook salmon smolts outmigrating through the lower portions of the Sacramento river system. Coded-wire-tagged (CWT) chinook salmon smolts were released at various locations within the river system between the years 1979 and 1995. Recoveries of these juvenile salmon in a lower river trawl fishery and later recoveries of adults from samples of ocean catches provided the basic data. Variation in recoveries, over and above differences in release numbers and fishing effort, was modeled with a generalized linear model having many covariates. The most influential covariates included environmental factors largely outside the control of man such as water temperature, water salinity, and river flow. Two covariates, more readily manipulated, were of particular interest: the position of a water diversion gate (open or closed) separating the mainstem from the central delta and the relative fraction of water exported for irrigation and urban consumption. Of these two, only gate position suggested a strong effect. When the gate was open, fish released upstream of the gate suffered increased mortality but survival increased for fish released in the central delta region (on the other side of the gate). Over the range of export levels observed, there was no strong evidence for either adverse or beneficial effects of increasing water exports.

Contents

List of Figures

List of Tables

1 Introduction and Summary

This document describes a statistical model that relates salmon smolt mortality in the Sacramento river delta to a number of potentially influential factors. The consequent summary of statistical evidence is intended to aid biologists and decision makers in understanding the complexities of salmon smolt survival and ultimately in developing water use policies that balance environmental, agricultural, urban and other interests. It is our intention that the model be used in conjunction with, and as a supplement to, other sources of knowledge and not in a prescriptive, "stand-alone" mode.

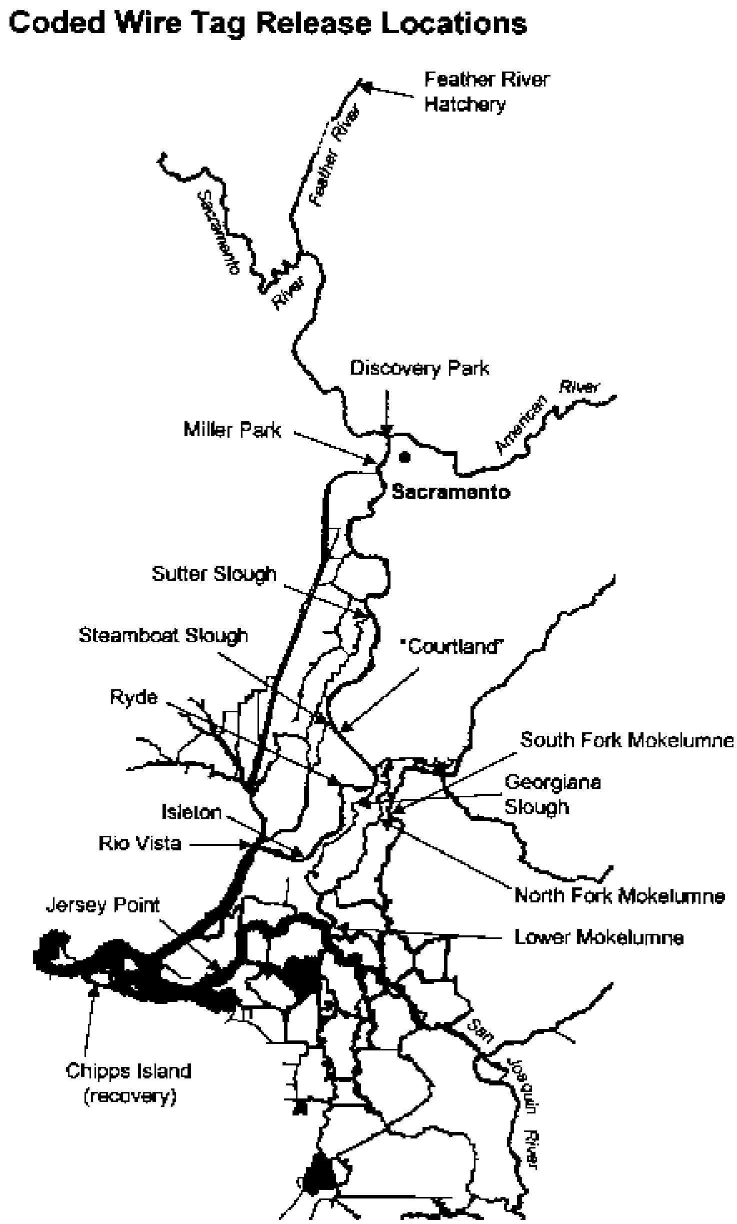
The model has been developed by a group representing fisheries interests, environmental concerns, government agencies, and water users. This effort was led by the authors with substantial input from Pat Brandes (US Fish and Wildlife Service) and Phyllis Fox (California Urban Water Agencies (CUWA)). Other key individuals were Elaine Archibald (consultant to CUWA), Randy Bailey (consultant to Metropolitan Water District), Jim Buell (consultant to Metropolitan Water District), Wim Kimmerer (Romberg Tiburon Center of San Francisco State University), and Tom Taylor (consultant to City of San Francisco). Numerous others provided suggestions and advice, including Alan Baracco, Sheila Greene, Martin Kjelson, Sharon Kramer, Dudley Reisser, Pete Rhoads, Terry Speed, and John Williams.

Reservations about an earlier model for salmon smolt survival (Kjelson, Greene, and Brandes, 1989) provided a primary impetus for the current effort. The earlier model only used river temperature, exports, and percent of flow diverted through the cross channel gates and Georgiana Slough as variables; the current model uses the more extensive set listed below. The current model is founded on an arguably more sensible probability structure leading to rather different statistical methodology. The new model incorporates the information contained in ocean recoveries in addition to Chipps Island recoveries, whereas the previous model only used the latter. The newer model used more releases—through 1995 as compared to through 1989 in the earlier model.

The data used in fitting the model consist of (1) recoveries from trawls taken at a downstream location, Chipps Island, and (2) later captures in the ocean fisheries of smolts released at a number of upstream locations in the Sacramento-San Joaquin system, primarily on the Sacramento side. The releases took place from 1979 to 1995. For each release group the values of a number of possibly influential factors, termed "covariates," were recorded. These included: release site, smolt size, Sacramento flow, salinity, annual quantities of rice pesticides applied, river water temperature, hatchery temperature, temperature shock (the temperature difference between water in transport truck and water in the river at the release site), a tidal variable, the ratio of water exports to river flow, position of the Delta cross channel gate, a time trend, and turbidity. The number of smolts released, a measure of the trawling effort at Chipps Island, and a measure of sampling effort in the ocean fisheries were also used. These data are described more precisely in Appendix A of the report. Figure 1 shows the release locations as well as the location of Chipps Island.

The statistical model rests on certain assumptions:

Figure 1: Release locations in lower Sacramento river system.



A : For a given release group we assume that the expected number of recoveries at Chipps Island is proportional to the product of: (1) the number of smolts released, (2) the probability of a smolt surviving from the release point to Chipps Island, and (3) the reported effort at Chipps Island. The constant of proportionality is an unknown "catchability coefficient" which is assumed to be independent of the covariates mentioned above and constant for all release groups.

B : For such a release group, we assume that the expected number of fish recaptured in the ocean fisheries at ages 2, 3, and 4 is equal to the product of: (1) the number released, (2) the probability of survival to Chipps Island, (3) the probability that a smolt surviving to Chipps Island is later caught in the ocean and recovered in a catch sample, and (4) an expansion factor that reflects catch sampling effort. It is assumed that (3) is constant for all groups released within the same year.

C : Recoveries are related to covariates through a model in which the logarithm of the expected number of recoveries is a linear function of covariate values, with coefficients to be determined by fitting.

D : The statistical variance of the number recovered is proportional to the expected number recovered.

E : The coefficients of some covariates are assumed to be the same for all release sites and those of others are allowed to differ from release site to release site; i.e., there is a release site interaction with these covariates.

F : We assume that the relationships between the covariates and mortality have not changed during the period of time for which the model is fitted.

These assumptions are presented in more detail in section 2. Sensitivity analysis presented in Section 5 provides some checks into these assumptions and of other aspects of model formulation and fitting.

The model directly estimates expected recovery at Chipps Island as a function of the covariates. Without making stronger assumptions, absolute survival probabilities to Chipps Island cannot be estimated. This is a consequence of the unknown catchability coefficient at Chipps Island. However, relative survival probabilities can be estimated. As examples, we can estimate the ratio of survival probabilities for releases from Ryde and Courtland with common values of the covariates, and we can estimate the ratio of survival probabilities for releases from Sacramento for different values of the covariates.

It is imperative to keep certain inevitable limitations in mind when interpreting results of the model:

- The model is undoubtedly a highly idealized approximation to a complex physical and biological reality. It is our hope that some of the important facets of this reality will be reflected in the results of the model, but there are no guarantees.

- The model results are based on correlations, and correlation does not imply causation.

- The predictive ability of the model depends on factors not included in the model staying similar to their status during the historical period of time in which the data were collected.
- Predictive ability outside the range of data observed is unknown.
- The coefficients of variables in the model cannot be viewed in isolation. They depend upon other variables in the model and upon unmeasured variables that are not in the model. A variable in the model may act as a proxy for variables outside the model; an example is the time trend variable.

We hope that qualitative information will be drawn from the model and used to supplement current biological understanding and to provide directions for further research. We envision that the primary quantitative use of the model will be to predict relative survival rates under different scenarios along with measures of the uncertainties of those estimates. Since low recovery rates produce substantial noise in the data, it is important to acknowledge the degree of uncertainty in any quantitative estimate. An example, developed in more detail in the section 4.2, compares the estimated ratio of survival probabilities under two strategies for a release from Sacramento. Strategy I consists of opening the cross channel gate and the export/inflow ratio is 0.4 and strategy II consists of closing the gate and an export/inflow ratio of 0.2. We estimate the relative survival under strategy I to that under second strategy to be 0.67 with a standard error of 0.08.

Such predictions are formed by linear combinations of covariate values. These predictions are more trustworthy than the coefficients of individual variables. As we warned above, the coefficients cannot be viewed in strict isolation; for example flow and salinity have a strong inverse relationship in the data used to fit the model. Similarly there are relationships between flow, release temperature, and hatchery temperature. Since there has been no deliberate controlled experimentation, it is impossible to untangle causal effects from the model alone. This said, we do cautiously attempt to draw some conclusions about the correlation of survival with individual factors or groups of factors:

- Release site makes a difference. Survival of releases from Sacramento, Sutter and Steamboat Sloughs, Courtland, and Ryde is higher than that of releases from Feather River Hatchery, Jersey Point, and Mokelumne and Georgiana Slough. Increased mortality in the central delta is consistent with earlier findings of the USFWS (Kjelson, et al., 1989).
- High release temperatures are associated with high mortality. Differences in temperature clearly distinguished releases with very high mortality from releases with very low mortality. This is consistent with the earlier model.
- The combined effect of flow and salinity is significant. Both appear to be important, but we cannot untangle differential effects.
- The cross channel gate being open is associated with lower survival for releases on the Sacramento and increased survival for releases in the central delta. This is broadly consistent with findings of the earlier USFWS model and with the first item above.

- There is an increasing trend in mortality over time. This trend probably reflects the effects of numerous causes of general environmental degradation. There is a negative association with annual quantity of applied rice pesticides that does not appear to be entirely explainable by the time trend. It is of course entirely possible that applied rice pesticide is acting as a surrogate for other variables that have deleterious effects and were not included in the model.
- There is little evidence for the importance of exports. Although the earlier model used exports, their estimated effect was also small.
- There is little evidence for the importance of tidal phase.

The data contain a large component of random variability, and mechanisms influencing mortality may not have emerged from our analysis because corresponding effects did not stand out above the noise. Furthermore, important variables may not have been identified or measured. However, the combination of careful statistical modeling and expert guidance and advice has made it possible to draw certain broad conclusions that we hope contribute to the debate on wise use of our limited water resources.

2 Covariates and Model Structure

First we introduce some notation and terminology.

S, r, and f represent survival rate, recapture probability conditional on survival, and recovery effort level, respectively. The product Sr is referred to as a recovery rate. Subscripts C or O refer to Chipps Island or the ocean.

S_C = probability an upstream fish survives to Chipps Island

r_C = capture probability at Chipps Island, assuming that the fish is alive at that location

π_{CO} = probability that an upstream fish surviving to Chipps Island is later caught in the ocean and recovered in a catch sample (roughly $\equiv S_O r_O / f_O$)

f_C = proportion of space-time sampled at Chipps Island

f_O = average expansion factor for ocean recoveries (#estimated/#observed)

R = number of fish released

Y_C = number of fish recaptured at Chipps Island

Y_O = (estimated) number of fish recaptured in ocean fishery (at ages 2, 3, or 4)

x_i = covariate i

The assumptions can be written more concisely now. For a given release group of R fish, assumption **A** implies

$$\begin{aligned} \mathrm{E}[Y_C] &= RS_C r_C && (1) \\ &= RS_C f_C q_C && (2) \end{aligned}$$

where the catchability coefficient at Chipps Island, q_C, is assumed constant, but unknown, for all releases. From assumption **B**,

$$\begin{aligned} \mathrm{E}[Y_O] &= RS_C(1-r_C)\pi_{CO} f_O && (3) \\ &\approx RS_C \pi_{CO} f_O && (4) \end{aligned}$$

where within a given year π_{CO} is constant for all release groups. An implicit assumption in going from equation (3) to equation (4) is that r_C is so small that $1-r_C \approx 1$, in other words that only a small fraction of the fish surviving to Chipps Island are captured there.

Assumption **D** can be expressed

$$\begin{aligned} \mathrm{Var}[Y_C] &= \phi_C \mathrm{E}[Y_C] && (5) \\ \mathrm{Var}[Y_O] &= \phi_O \mathrm{E}[Y_O] && (6) \end{aligned}$$

The parameters ϕ_C and ϕ_O are called dispersion parameters[1].

The modeling of $\mathrm{E}[Y_C]$ and $\mathrm{E}[Y_O]$ as functions of covariates is based on Assumption **C**. Release size R is treated as a known constant as are f_C and f_O. Therefore, from equations (2) and (4), using p covariates for Chipps Island recoveries and s covariates for ocean recoveries

$$\begin{aligned} \log(E[Y_C]) &= \log(Rf_C) + \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p \\ \log(E[Y_O]) &= \log(Rf_O) + \beta'_0 + \beta'_1 x'_1 + \beta'_2 x'_2 + \ldots + \beta'_s x'_s \end{aligned}$$

Some of the covariates are shared by both recovery locations as are some of the coefficients and the relationships between β and β' are given below.

Thus the logarithms of q_C, S_C, and π_{CO} are being modeled as functions of covariates. Primary interest is in S_C, which is assumed to be a function of two types of covariates: (1) factors whose influences are independent of release site, and (2) release site dependent factors. The effects of the site dependent covariates, export levels, gate position, and turbidity, were allowed to vary between groups of sites; i.e., an interaction between release site and these three covariates was modeled. For example, the coefficient for export level could differ between fish released at Courtland and fish released at Mokelumne.

The covariates are categorized below, with variable name given in italics (see Appendix A for detailed descriptions). An explanation of the variable selection procedure follows.

Site Independent (SI) :

1. *Size* = fish size (mm)

[1] If Y_C and Y_O were Poisson random variables, then ϕ would be 1.0.

2. *log.Sacramento.2* = natural logarithm of Sacramento river flow (cfs)
3. *Collinsville* = salinity measured at Collinsville (micro mho/cm)
4. *Pesticide* = annual applied rice pesticide (pounds)
5. *Trend* = linear annual trend
6. *Release.Temp* = release temperature (degrees Fahrenheit)
7. *Hatch.Temp* = hatchery temperature (degrees Fahrenheit)
8. *Shock* = release temperature - truck temperature (degrees Fahrenheit)
9. *Tide.Var* = early tidal asymmetry factor plus early tidal trend factor (feet)

Site Dependent (SD) :

1. Release Location: Release sites were grouped into the following seven categories[2]:
   - *frh.dum* = Feather River Hatchery
   - *sac.dum* = Discovery Park, Miller Park, and Sacramento
   - *slo.dum* = Steamboat and Sutter Sloughs
   - *crt.dum* = Courtland
   - *ryd.dum* = Rio Vista, Isleton, and Ryde
   - *mkg.dum* = Lower Mokelumne, North Fork Mokelumne, South Fork Mokelumne, and Georgiana Slough
   - Jersey Point
2. Exports/Inflow Ratio (cfs/cfs)
   - *upper.exp.inflow* = Feather River Hatchery, Discovery Park, Miller Park, Sacramento, Courtland
   - *delta.exp.inflow* = Lower Mokelumne, North Fork Mokelumne, South Fork Mokelumne, Georgianna Slough, and Jersey Point
3. (Cross-channel) Gate Position: coded as 1 for open and 0 for closed;
   - *upper.gate* = Feather River Hatchery, Discovery Park, Miller Park, Sacramento, Courtland
   - *delta.gate* = Lower Mokelumne, North Fork Mokelumne, South Fork Mokelumne, Georgianna Slough, and Jersey Point
4. Turbidity (Formazine turbidity units)
   - *mainstem.turbid* = mainstem turbidity value used for Feather River Hatchery, Discovery Park, Miller Park, Sacramento, Courtland, and Ryde (including Rio Vista and Isleton)

[2] The first six site groupings had corresponding dummy variables. A dummy variable takes on the value 1.0 when the "effect" is present. For example, *frh.dum* equals 1.0 when the release location is Feather River Hatchery and 0 for releases made elsewhere. For a release made at Jersey Point all site dummy variables are set equal to zero. The effect of dummy variables is to shift the intercept up or down. When a dummy variable is multiplied times a continuous covariate, like export/inflow ratio, say, thus creating a new variable, the slope of the model can differ between cases with the "effect" and without the "effect".

- *delta.turbid* = central delta turbidity value used for Lower Mokelumne, North Fork Mokelumne, South Fork Mokelumne, Georgianna Slough, and Jersey Point

The variable coding for the covariate groupings of exports/inflow ratio, gate position, and turbidity coefficients was a product of a dummy variable for the release site and the covariate value. For example, a release group from Sacramento would use the observed export/inflow ratio value for that release as *upper.exp.inflow* and a zero for *delta.exp.inflow*. If the gate were open for that same release group *upper.gate* would equal 1.0 and *delta.gate* would equal zero. Finally the mainstem turbidity value would be used for *mainstem.turbid* and a zero would be used for *delta.turbid*.

Note that releases from Steamboat Slough or Sutter Slough are assumed unaffected by the export/inflow ratio, cross-channel gate position, and mainstem (or central delta) turbidity. Releases at these locations are geographically removed from the primary water pumping locations. The fish are unlikely to travel upstream into the mainstem and thus be affected by the gate position. The sloughs empty back into the mainstem about midway between the release points on the sloughs and Chipps Island, so for survivors that do reach the mainstem, we are assuming the effect of variations in mainstem turbidity is relatively negligible.

Similarly releases from Ryde are assumed unaffected by the export/inflow ratio and gate position, but given its location on the mainstem, mainstem turbidity could have some effect. Whatever effect there may be is assumed identical for all other mainstem releases.

Other covariates included a dummy variable for Chipps Island recoveries (the intercept could differ between Chipps Island recoveries and ocean recoveries), and, for ocean recoveries, dummy variables for differing year effects (*1979 1980 ... 1993*) with 1994 the default year[3]. The effect of the catchability coefficient at Chipps Island, q_C, is partially modeled through the Chipps Island dummy variable. The ocean survival-capture combination π_{CO} for each year is also modeled, to some degree, by the year effect variables; i.e., the intercept could shift up or down between years.

The model formulation for the expected Chipps Island and ocean recoveries follows from (2) and (4) and the above covariates (symbolically):

$$\begin{aligned}
\log(E[Y_C]) &= \log(Rf_C) + \log(q_C) + \log(S_C) && (7)\\
&= \log(Rf_C) + \beta_0 + \beta_1\text{Chipps Dummy} + \beta_2\text{SI} + \beta_3\text{SD} && (8)\\
\log(E[Y_O]) &= \log(Rf_O) + \log(S_C) + \log(\pi_{CO}) && (9)\\
&= \log(Rf_O) + \beta_0 + \beta_2\text{SI} + \beta_3\text{SD} + \beta_4\text{Year Dummies} && (10)
\end{aligned}$$

where β_0 is a common intercept.

[3]For releases made during 1994, the covariates for all the dummy year coefficients *1979 1980 ... 1993* are set equal to zero. As a result the values of the coefficients for years other than 1994 are relative to 1994 releases, with negative coefficients indicating a year worse than 1994 and positive coefficients indicating a year better than 1994. There is no indicator for 1995 releases because those releases did not have ocean recoveries at the time the data set was created

2.1 Covariate selection

Selection of covariates was done largely before model fitting, but after examination of informative bi-variate plots and group discussion. The strategy was to include most covariates having an *a priori* reasonable possibility of being influential while eliminating covariates that were somewhat highly correlated. For example, of two highly correlated measures of salinity, only one was used for modeling.

After an initial model was fit, however, additional changes were made. In particular the flow rate variable was log transformed after a discussion of the possible multiplicative effect of flow on survival. After observing an upward trend in pesticide levels, a simple trend variable was added to partially separate long term changes in the environment from changes in pesticide level. A measure of pesticide concentration was tried instead of total pounds of pesticide as well, but the quality of fit worsened considerably. Out of concern over potential temperature shock after fish leave the transport vehicle and enter the river, a new covariate, Shock or Release Temperature minus Truck Temperature, was created and substituted for Truck Temperature. Site specific effects from gate position and export levels were originally allowed for Ryde releases, even though Ryde is located below the gate, because in principle the tide could push the smolt back upstream. Unrealistic coefficient values, however, led to the removal of these Ryde-specific covariates. Finally, two competing measures of water exports were compared, the export/inflow ratio (the measure selected and discussed previously) and total exports. The denominator of the export/inflow ratio is very highly correlated with Sacramento flow— thus to minimize problems with collinearity between covariates only two of the following three covariates, Sacramento flow, export/inflow ratio, and exports, should be in a model. Export/inflow ratio is currently a popular measure of export level, and since substitution of export/inflow ratio for exports led to no sizeable changes in goodness of fit nor estimates of other coefficients (see section on sensitivity analyses), export/inflow ratio was chosen for inclusion.

2.2 Estimation of relative survival rates

The assumptions made allow estimation of the *absolute recovery rate to Chipps Island*, $S_C r_C$, for any given release group as well as estimation of the *relative survival rates* to Chipps Island, say $S_{C,1}/S_{C,2}$, for any two release groups. If these parameters can be estimated, they can be modeled as functions of covariates. The reader not interested in technical details may want to skip this and the following subsection.

For instructional purposes, we first demonstrate that the *absolute recovery rate to Chipps Island* can be estimated from the 'raw' recoveries. This approach is done independently of the covariate-based model and the estimate will be referred to as a *raw* estimate. From (2), for a given release group *absolute recovery rate to Chipps Island*, $S_C r_C$, can be estimated by:

$$\widehat{S_C r_C} = \frac{Y_C}{R} \tag{11}$$

The *absolute survival rate to Chipps Island*, S_C, however, cannot be separately estimated, because $r_C = q_C f_C$ and q_C is unknown. Because primary interest is in S_C, estimating the absolute recovery rate $S_C r_C$ is not that useful. Relative survival rates, however, can be estimated and provide a means of comparing release strategies.

The *relative survival rates* to Chipps Island can be estimated by dividing both sides of (11) by $f_{C,1}$ or $f_{C,2}$ and taking the ratio of the resulting estimates $S_{C_1} q_C$ and $S_{C_2} q_C$, where the q_C's will cancel.

$$\widehat{\frac{S_{C,1}}{S_{C,2}}} = \frac{Y_{C,1}/(R_1 f_{C,1})}{Y_{C,2}/(R_2 f_{C,2})} \tag{12}$$

The relative survival rates can also be estimated from the ocean recoveries.

$$\widehat{\frac{S_{C,1}}{S_{C,2}}} = \frac{Y_{O,1}/(R_1 f_{O,1})}{Y_{O,2}/(R_2 f_{O,2})} \tag{13}$$

Thus two *raw* estimates of the same quantity, (12) and (13), can be calculated. Combining the two estimates in a weighted average would be sensible, where the weights were reflective of variances for each estimates (and the covariance between).

2.3 Covariate-based versus *raw* estimates of relative survival rates

Here we show that the covariate-based model (equations (8) and (10)) has the same limits as to what can be estimated. Consider two groups with identical values of R released at the same time and exposed to identical effort levels, f_C, at Chipps Island (if not the f_C's could be divided out) but differing covariates. From (7),

$$\begin{aligned}
\log(E[Y_{C,1}]) - \log(E[Y_{C,2}]) &= \log(R_1 f_{C,1} q_{C,1}) + \log(S_{C,1}) - \log(R_2 f_{C,2} q_{C,2}) - \log(S_{C,2}) && (14)\\
&= \log(S_{C,1}) - \log(S_{C,2}) && (15)\\
&= \beta_2(SI_1 - SI_2) + \beta_3(SD_1 - SD_2) && (16)
\end{aligned}$$

So,

$$\frac{S_{C,1}}{S_{C,2}} = \exp\left(\beta_2(SI_1 - SI_2) + \beta_3(SD_1 - SD_2)\right) \tag{17}$$

Thus relative survival rates are determined by the coefficients β_2 and β_3 which can be estimated from the data.

Contingent on the assumptions holding (particularly assumption **C**), the model-based approach has several advantages over the raw estimates. The first three advantages are analogous to those of a simple linear regression model, $\mathrm{E}[y] = \beta_0 + \beta_1\ x$, over a simple historical average $\overline{y}$. First, model-based estimates of relative survival rates are likely to be more precise, have a smaller variance, because a larger sample size has been effectively used by incorporating information from all release groups simultaneously. Second, a covariate-based model facilitates estimates or predictions for future release groups with covariates values slightly differing from those in the historical data set. Third, a covariate-based model, once fitted, can be studied to obtain information about the relative influence of different covariates on the response variable.

A fourth advantage is that the covariate-based model, as formulated here, conveniently utilizes information provided by both Chipps Island and ocean recoveries variables simultaneously as opposed to determining weights for combining (12) and (13). Information about the coefficients, 'β_2' and 'β_3', for the SI and SD covariates provided by ocean recoveries is implicitly incorporated into the parameter estimates.

Conversely, the disadvantage of the covariate-based approach is the requirement that additional assumptions hold. If the assumptions do not hold, then bias is introduced, just as in the case of simple linear regression when in fact $\mathrm{E}[y|x] \neq \beta_0 + \beta_1 x$.

3 Model Fitting

The final formulation of the model described by equations (2) and (4) required estimating $p = 38$ coefficients corresponding to the covariates and two dispersion parameters (for Chipps Island recoveries and for ocean recoveries). The detailed model structure is shown in equation (18).

$$
\begin{aligned}
\log(E[Y]) = {} & \log(Rf) + \beta_0 + \beta_1 chipps.dum + \beta_2 Size + \beta_3 Log.Sacramento.2 \\
& + \beta_4 Collinsville + \beta_5 Pesticide + \beta_6 Trend + \beta_7 Release.Temp \\
& + \beta_8 Hatchery.Temp + \beta_9 Shock + \beta_{10} Tide.Var \\
& + \beta_{11} frh.dum + \beta_{12} sac.dum + \beta_{13} slo.dum + \beta_{14} crt.dum \\
& + \beta_{15} ryd.dum + \beta_{16} mkg.dum + \beta_{17} upper.exp.inflow + \beta_{18} delta.exp.inflow \\
& + \beta_{19} upper.gate + \beta_{20} delta.gate + \beta_{21} mainstem.turbid \\
& + \beta_{22} delta.turbid + \beta_{23} 1979.dum + \ldots + \beta_{37} 1993.dum
\end{aligned} \tag{18}
$$

For the 86 Chipps Island observations the indicator variable *chipps.dum* equalled one and for the 84 ocean observations equalled zero[4]. The year indicator variables, 1979.*dum*, ..., 1993.*dum*, were all set equal to 0 for the Chipps Island observations.

The covariates were standardized before estimating the model coefficients. For example, for *Size*, the average fish length for the 170 observations was 80.01 mm with a standard deviation of 6.47 mm; so 80.01 mm was subtracted from each value and the result divided by 6.47 mm. This has the effect of minimizing numerical errors on the computer as well as making the estimated coefficients somewhat more comparable. The coefficient of a particular variable represents the change in fitted log recovery as a function of that variable measured in units of standard deviations rather than in the original units. We discuss the interpretation of these coefficients further in Section 4. The means and standard deviations for all the covariates are found in Appendix C.

The coefficients, $\beta_0, \ldots, \beta_{37}$, were estimated using iterated weighted least squares (IWLS)

[4]Sample sizes differ for the two types of observations because ocean recovery information was not yet available for the 1995 releases, one of the 1994 releases had no recorded age 2 recoveries (thus the ocean catch sample expansion factor could not be calculated), and one of the Chipps Island observations was deleted as an outlier (see Appendix A).

(pp. 40–43, McCullagh and Nelder, 1989) with a ridge regression parameter, λ (pp. 255–257, Weisberg, 1985). Ignoring the ridge aspect momentarily, the estimated coefficients correspond to those arising from maximizing a quasi-likelihood function. A quasi-likelihood function is based purely on mean and variance (Chapter 9, McCullagh and Nelder, 1989) of the probability distribution for the observations and is an approach to dealing with situations where one does not want to or cannot accurately specify the distribution in its entirety. If the dispersion parameters had been fixed at 1.0 for both Y_C and Y_O, then the equality of the mean and the variance (equations (5) and (6)) would imply a Poisson distribution.

The reason for the ridge parameter is the large number of coefficients to estimate, 38, relative to the number of observations, 170. When the ratio of coefficients to observations is relatively large, the variances of the estimated coefficients tend to be large. The ridge regression approach was used as a means of reducing these variances; given fewer coefficients this approach would not be necessary. The ridge parameter was set at λ=550, where the value was selected partially by examining the leave-one-out cross-validation prediction errors over a wide range of λ values.

The dispersion weights for Chipps Island and ocean recoveries, ϕ_C and ϕ_O, were estimated simultaneously in an iterative manner based on squared residuals.

Standard errors for the estimated coefficients were estimated using an analytical approximation. The approximation ignores possible time dependence between observations, for example, the correlation of environmental conditions in adjacent years, and as such may yield slight underestimates. The standard errors also ignore the data-based choice of the ridge parameter λ.

Further technical details of the estimation procedures are sketched in Appendix B.

4 Results

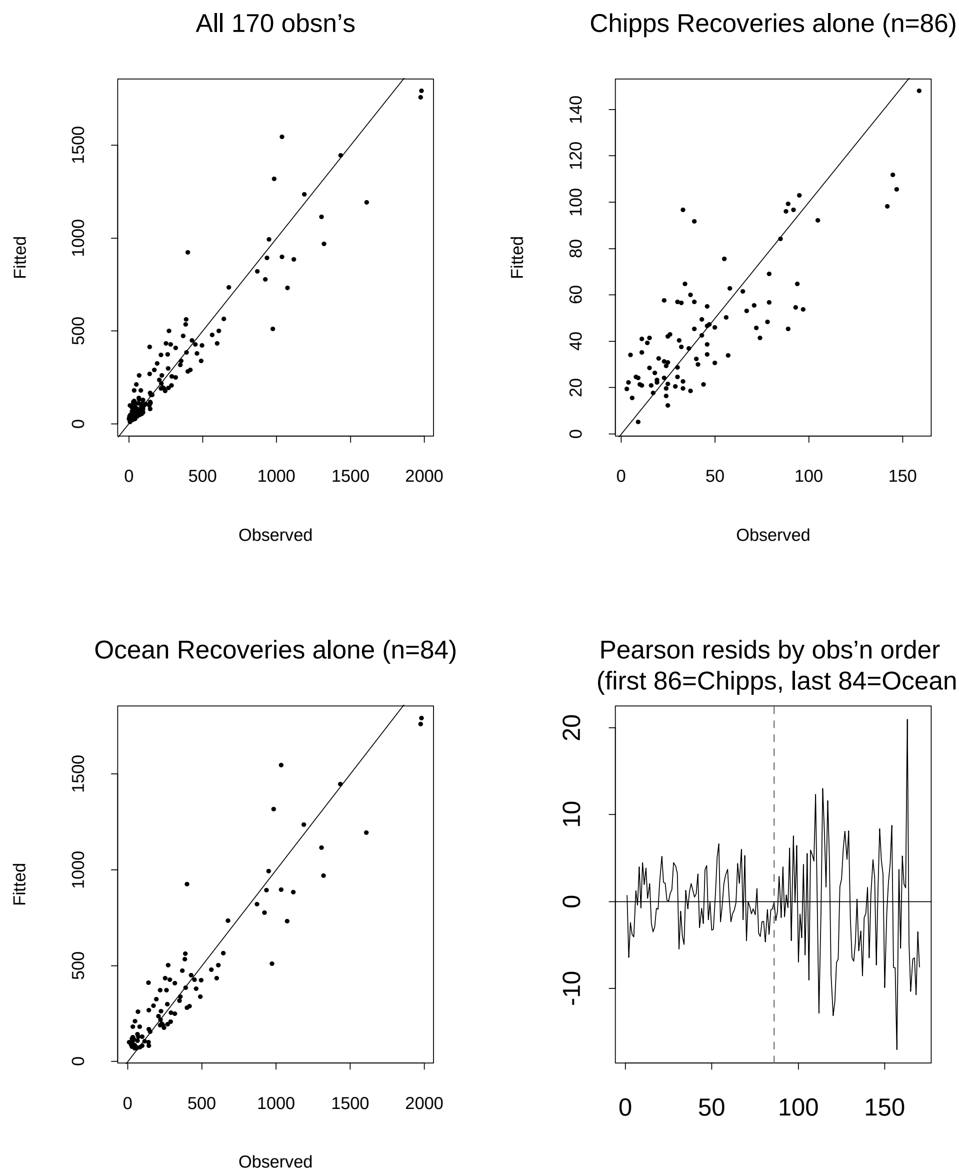
Estimates of the coefficients, β_0, ..., β_{37}, and their standard errors are given in Table 1. The estimated dispersion parameters were $\hat{\phi}_C$=15.26 and $\hat{\phi}_O$=84.55. Recall that the Site Dependent coefficients corresponding to the export, gate position, and turbidity covariates are release site specific. For example, the site dependent effects for a Feather River Hatchery release are reflected by *frh.dum*, *upper.exp.inflow*, *upper.gate*, and *mainstem.turbid.*

Figure 2 includes plots of the fitted versus observed recoveries. The upper left plot is for all 170 observations while the upper right and lower left are plots for the Chipps Island recovery subset (n = 86) and the ocean recovery subset (n = 84). The increasing variation of fitted values as observed values increase is consistent with the assumption that variance is proportional to expected value. The lower right plot is of the Pearson residuals, $(y - \hat{y})/\sqrt{\hat{y}}$, a scaling of the residuals that adjusts for differences in variances. The x-axis is the 170 observations ordered with the first 86 being the Chipps Island recoveries and the last 84 being the ocean recoveries. The difference in the magnitude of the residuals is evidence of the differences in the dispersion parameter between the two recovery sets. The observed and fitted values are given in Appendix D.

Table 1: Estimated coefficients and standard errors (se). $\hat{\beta}_\lambda$ is the coefficient for standardized covariates and $\hat{\beta}^*_\lambda$ is for unstandardized covariates. Default site location is Jersey Point and default release year is 1994.

Covariate	$\hat{\beta}_\lambda$	se($\hat{\beta}_\lambda$)	$\hat{\beta}^*_\lambda$
Intercept	-5.982	0.046	-2.92e+00
Chipps	1.026	0.071	2.05e+00
Site Independent Factors			
Size	0.076	0.051	1.17e-02
Log.Sacramento.2	0.161	0.064	3.15e-01
Collinsville	0.278	0.062	7.24e-05
Pesticide	-0.173	0.055	-2.67e-07
Trend	-0.069	0.065	-1.79e-02
Release.Temp	-0.285	0.063	-6.34e-02
Hatchery.Temp	-0.090	0.063	-2.92e-02
Shock	-0.049	0.060	-9.52e-03
Tide.Var	-0.028	0.046	-5.70e-02
Site Dependent Factors			
frh.dum	-0.190	0.048	-1.25e+00
sac.dum	0.044	0.052	1.09e-01
slo.dum	0.071	0.049	2.99e-01
crt.dum	0.043	0.050	1.22e-01
ryd.dum	0.147	0.053	3.27e-01
mkg.dum	-0.126	0.060	-3.29e-01
upper.exp.inflow	-0.037	0.063	-2.25e-01
delta.exp.inflow	-0.070	0.068	-4.90e-01
upper.gate	-0.145	0.052	-3.50e-01
delta.gate	0.131	0.068	3.35e-01
mainstem.turbid	-0.047	0.059	-6.78e-03
delta.turbid	-0.036	0.067	-7.47e-03
Year Effects			
1979	-0.103	0.057	-1.34e+00
1980	0.087	0.041	5.75e-01
1981	-0.059	0.070	-7.63e-01
1982	0.062	0.038	8.14e-01
1983	-0.069	0.042	-5.25e-01
1984	0.127	0.049	7.49e-01
1985	0.049	0.036	3.24e-01
1986	0.237	0.033	1.79e+00
1987	0.243	0.029	1.60e+00
1988	0.247	0.043	1.00e+00
1989	-0.040	0.059	-1.46e-01
1990	-0.035	0.049	-1.75e-01
1991	-0.011	0.046	-6.17e-02
1992	0.026	0.048	1.42e-01
1993	0.239	0.042	1.20e+00

Figure 2: Fitted versus observed recoveries plus residual plot.



The estimated coefficients for the SI and SD covariates are plotted in descending order in Figure 3 along with $\pm$ 2 standard errors. This provides an approximate means of visually separating strong from weak effects in that coefficients with these approximate confidence intervals including zero would be considered weak (namely the true coefficient could be zero).

4.1 Using the model for prediction

The recommended use of the model for prediction is comparing the effect of different release strategies on survival to Chipps Island as opposed to predicting absolute recoveries either at Chipps Island or in the ocean. In particular, we recommend that the ratio of recovery rates at Chipps Island be estimated rather than differences in recovery numbers.

To make estimates of the absolute number of Chipps Island recoveries for a given release group, one needs estimates of the trend term, which will be problematic given that future releases will be made in years outside the data base.

Similarly, to make estimates of the absolute number of ocean recoveries requires the trend value as well as a guess as to what previous year in the data base will most resemble the release year to be forecast.

On the other hand, when comparing the effect of different release strategies, comparisons will (typically) be between releases made the same year. The unknown trend effects will be the same in that case, thus cancelling when the ratio of recoveries is calculated. The ratio of ocean recovery rates will be the same as the ratio of Chipps Island recovery rates, as well, since the ocean year effect covariates (whatever they are) will cancel. Thus the emphasis is on the relative impact on survival rates through the delta of one strategy versus another.

An example is given in the next section, but first we show how to estimate one of the components in a ratio (the term that will be either the numerator or denominator with respect to Chipps Island recoveries). Calculations can be made in terms of covariates as measured on the original scale or standardized. The coefficients in the third column of Table 1 are for the covariates rescaled to original units and are denoted $\hat{\beta}^*_\lambda$. The log of the expected number of returns was modeled, so expected number of returns is

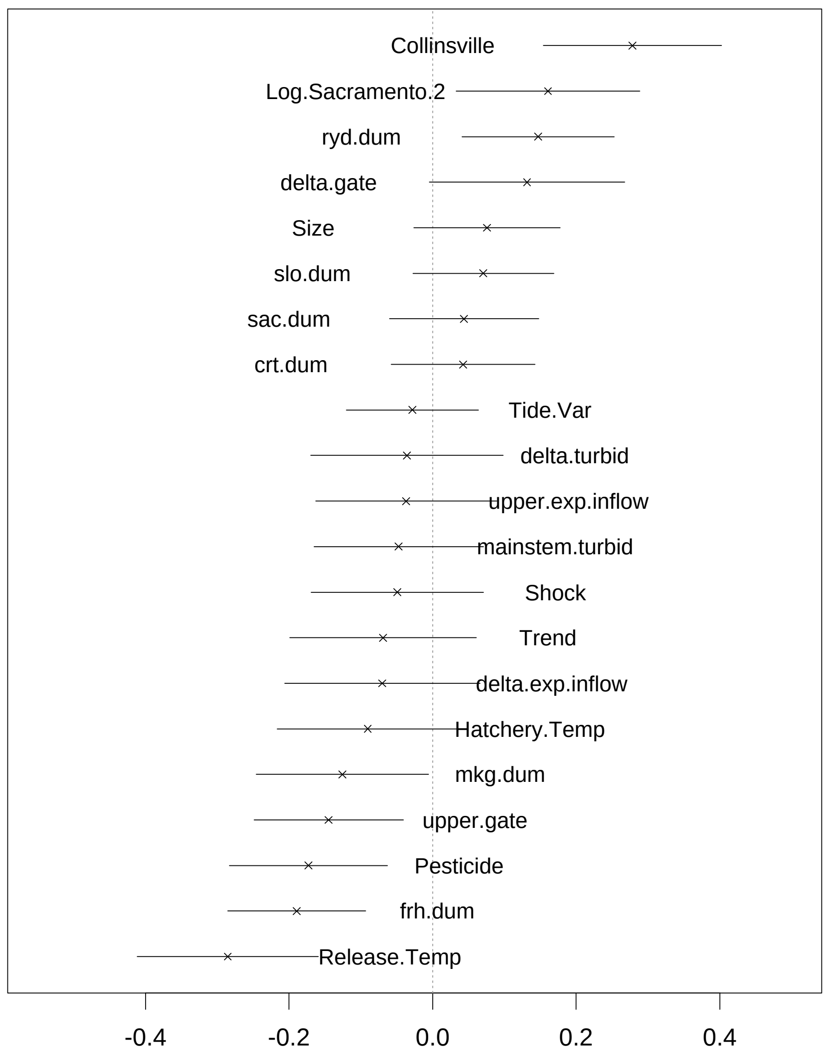
$$Rf_C \exp\left[\hat{\beta}^*_{\lambda,0} + \sum_{i=1}^{37} \hat{\beta}^*_{\lambda,i} x_i\right], \tag{19}$$

where x_i is the value of the ith covariate. Alternatively the covariates can be standardized and $\hat{\beta}_\lambda$ (column 1 of Table 1) used instead. The means and standard deviations needed to standardize each covariate are given in Appendix C. The expected number of returns is calculated by

$$Rf_C \exp\left[\hat{\beta}_{\lambda,0} + \sum_{i=1}^{37} \hat{\beta}_{\lambda,i} \frac{x_i - \overline{x}_i}{s_{x_i}}\right], \tag{20}$$

where $\overline{x}_i$ and s_{x_i} are the mean and standard deviation for the ith covariate. It is much simpler to calculate standard errors when the covariates are standardized (see Appendix B).

Figure 3: Estimated coefficients for SI and SD covariates $\pm$ 2 se's.



A general caveat is that using combinations of covariates outside the range of the data set (see Appendix A) may provide misleading results. This is addressed further in the Discussion.

For example, suppose 50,000 fish are to be released from Sacramento and the recovery effort, f_C, will be 0.10. SI covariate values are Size= 78 mm, Sacramento flow = 18,000 cfs, Collinsville salinity = 2,000 micro mhos/cm, pesticide = 3,000,000 pounds, release temperature = 62° F, hatchery temperature = 56° F, shock = 9° F, and tide variable = 1.9. The trend value will be based on a 1994 year. The gate will be open, export/inflow is 0.4, and the mainstem turbidity is 7.0 FTUs. Recall that the covariate values for the dummy or indicator variates are set equal to 1 when the 'effect' is true[5]. The coding for the gate position is 1 when open and 0 when closed. So $sac.dum$=1 and $upper.gate$=1, while $frh.dum = slo.dum = crt.dum = mkg.dum = delta.gate = 0$. Similarly $upper.exp.inflow$=0.4 while $delta.exp.inflow$=0 and $mainstem.turbid$=7.0 with $delta.turbid$=0. The number of recoveries at Chipps Island is estimated to be 23.9[6] using

$$\begin{aligned}
\hat{y} &= 50000 \times 0.10 \times \exp[-2.92 + (2.05 \times 1)[Chipps] + (0.0117 \times 78)[Size] + (0.315 \times \log(18000))[Log.Sacramento.2] \\
&+ (0.0000724 \times 2000)[Collinsville] + (-0.000000267 \times 3,000,000)[Pesticide] \\
&+ (-0.0179 \times 94)[Trend] + (-0.0634 \times 62)[Release.Temp] + (-0.0292 \times 56)[Hatchery.Temp] \\
&+ (-0.00952 \times 9)[Shock] + (-0.0570 \times 1.9)[Tide.Var] + (0.109 \times 1)[sac.dum] \\
&+ (-0.225 \times 0.4)[upper.exp.inflow] + (-0.350 \times 1)[upper.gate] + (-0.00678 \times 7.0)[mainstem.turbid] \qquad (21)
\end{aligned}$$

Note that the covariates with value 0 need not be included in this calculation.

The standard error for the expected, or average, value (based on equation (25)) is 5.02 and the standard error for a predicted value is 19.76 (based on equation (26)).

We will partially demonstrate the standard error calculations for the expected value first. First standardize *all* the covariates, including those with 0 values; for example, $x_1^* = (1-0.5059)/0.5014$, for $Chipps$ Indicator, $x_2^* = (78 - 80.01)/6.469$ for $Size$, ..., $x_{37}^* = (0 - 0.04118)/0.1993$ for 1993 indicator. For the intercept use x_0^*=1. Let $\mathbf{x}^*$ be the resulting 38 by 1 column vector, and $\mathbf{x}^{*t}$ be the 1 by 38 row vector. Pre-multiply the variance-covariance matrix for $\hat{\beta}_0$, $\hat{\beta}_1$, ..., $\hat{\beta}_37$, Var$[\hat{\beta}]$, a 38 by 38 matrix (see Appendix E), by the row vector $\mathbf{x}^{*t}$. Postmultiply the result by the column vector $\mathbf{x}^*$. The result for the above example is 0.043976. Finally take the square root of the result and multiply it by the predicted value; for example, 23.9. In short

$$\begin{aligned}
\hat{se}(\hat{\mathrm{E}}[Y]) &= \hat{\mathrm{E}}[Y]\sqrt{\mathbf{x}^{*t}\mathrm{Var}[\hat{\beta}]\mathbf{x}^*} \\
&= 23.9 \times \sqrt{0.043976} \\
&= 5.02
\end{aligned}$$

The standard error for the predicted value can be estimated by adding the square of $\hat{se}(\hat{\mathrm{E}}[Y])$ to the product of the predicted value and the dispersion parameter estimate. Then

[5]This is on the original, unstandardized scale; i.e., when using $\hat{\beta}_\lambda^*$.

[6]Due to rounding of coefficients in the expression below the calculation using these values is 23.8.

take the square root of the sum:

$$\begin{aligned} \hat{se}(\hat{Y}) &= \sqrt{\hat{se}(\hat{\mathrm{E}}[Y])^2 + \hat{\phi}_C \hat{\mathrm{E}}[Y]} \\ &= \sqrt{5.02^2 + 15.26 \times 23.9} \\ &= 19.76 \end{aligned}$$

4.2 Example: estimating relative survival rates

To estimate the ratio of two release groups' survival, one can, of course, calculate the estimated number of recoveries (assuming equal R and f_C) separately (as in the above example) and then divide one estimate by the other. However, because of cancellations, only those covariates for which the groups have differing values, and those coefficients that differ for identical covariates, need to be considered.

For example, suppose that interest is in comparing the survival of fish under differing gate positions and export/inflow ratios. The fish are to be released from Sacramento and all other covariates, release numbers, fish sizes, Sacramento flows, etc, are identical.

- Strategy I: the gate is open and export/inflow ratio is 0.4
- Strategy II: the gate is closed and export/inflow ratio is 0.2

The only relevant coefficients are those for $upper.exp.inflow$ and $upper.gate$, where the covariate for $upper.gate$ when the gate is closed is zero. The calculations can be eased by simply using the difference in covariate values as new covariates. The relative survival of the first strategy to the second strategy:

$$\begin{aligned} \hat{\mathrm{E}}\left[\frac{S_{\text{Strategy I}}}{S_{\text{Strategy II}}}\right] &= \exp((-0.225 \times (0.4 - 0.2))[upper.exp.inflow] + (-0.350 \times (1 - 0))[upper.gate] \\ &= 0.674 \end{aligned} \tag{22}$$

The estimated standard error in this case is 0.0838 and the estimated prediction error is 0.825 (see Appendix B.2). On average, for every 10 fish reaching Chipps Island under Strategy II, only 6.74 $\pm$ 2 $\times$ 0.838, or 5.06 to 8.42 (the latter a crude 95% confidence interval), should reach Chipps Island under Strategy I.

The standard error calculation is simplest when the covariates are standardized. Unlike the absolute recovery rate case, for the standard error all that needs to be considered are the covariates with differing values because of cancellations in the ratio. The standard deviations for $upper.exp.inflow$ and $upper.gate$ are 0.164 and 0.414 (Appendix C). As a check, the point estimate based on standardized covariates matches that based on unstandardized covariates:

$$\exp\left[\frac{0.4-0.2}{0.164}(-0.037) + \frac{1-0}{0.414}(-0.145)\right] = 0.674$$

To estimate the standard error, first calculate the vector of standardized differences, namely

$$\left[\frac{0.4-0.2}{0.164}\ \frac{1-0}{0.414}\right] = [1.22\ 2.42]$$

Next take the portion of the covariance matrix for the coefficients corresponding to *upper.exp.inflow* and *upper.gate* (Appendix E) and pre-multiply it by [1.22 2.42] and post-multiply it by the transpose, $[1.22\ \ 2.42]^t$:

$$[1.22\ 2.42] \begin{bmatrix} 0.00401 & -0.00102 \\ -0.00102 & 0.00266 \end{bmatrix} \begin{bmatrix} 1.22 \\ 2.42 \end{bmatrix}$$

= 0.01549. The standard error is the product of point estimate and the square root of 0.01549:

$$\begin{aligned} \hat{se}\left(\hat{\mathrm{E}}\left[\frac{S_{\text{Strategy I}}}{S_{\text{Strategy II}}}\right]\right) &= 0.674 \times \sqrt{0.01549} \\ &= 0.0839 \end{aligned}$$

The standard error for a predicted ratio is found by taking the square root of the sum of the estimated ratio and the standard error for the expected ratio squared:

$$\begin{aligned} \hat{se}\left(\frac{\hat{S}_{\text{Strategy I}}}{\hat{S}_{\text{Strategy II}}}\right) &= \sqrt{\frac{\hat{S}_{\text{Strategy I}}}{\hat{S}_{\text{Strategy II}}} + \hat{se}^2\left(\hat{\mathrm{E}}\left[\frac{S_{\text{Strategy I}}}{S_{\text{Strategy II}}}\right]\right)} \\ &= \sqrt{0.674 + 0.0839^2} \\ &= 0.825 \end{aligned}$$

5 Sensitivity Analyses

To evaluate the reasonableness of assumptions as well as some alternative formulations, several sensitivity analyses were conducted including

- Reasonableness of assumption that π_{CO} is constant within a release year
- Influence of SI covariates on estimated coefficients for other SI covariates
- Residual analyses.
- Influence of individual observations on estimated coefficients
- Effect of using the ocean observations compared to the Chipps Island observations
- Separate size coefficients for Chipps Island and ocean recoveries
- Effect of removing high flow (> 20,000 cfs) observations
- Effect of alternative measure of exports

Table 2: Estimated SI coefficients when a single SI covariate is omitted.

Covariate	Value with all SI covs	Omitted covariate								
		-Size	-Log.Sac	-Coll	-Pest	-Tren	-Rele	-Hatc	-Shoc	-Tide
Size	0.076	—	0.063	0.121	0.088	0.087	0.065	0.069	0.076	0.068
Log.Sacramento.2	0.161	0.143	—	0.033	0.147	0.172	0.172	0.155	0.158	0.159
Collinsville	0.278	0.292	0.207	—	0.282	0.260	0.262	0.266	0.274	0.281
Pesticide	-0.173	-0.172	-0.162	-0.186	—	-0.197	-0.166	-0.166	-0.183	-0.175
Trend	-0.069	-0.089	-0.092	0.008	-0.163	—	-0.100	-0.043	-0.059	-0.066
Release.Temp	-0.285	-0.276	-0.295	-0.292	-0.293	-0.295	—	-0.323	-0.319	-0.274
Hatchery.Temp	-0.090	-0.079	-0.080	-0.058	-0.072	-0.073	-0.192	—	-0.093	-0.095
Shock	-0.049	-0.050	-0.041	-0.018	-0.090	-0.037	-0.217	-0.053	—	-0.053
Tide.Var	-0.028	-0.012	-0.028	-0.049	-0.043	-0.026	0.013	-0.036	-0.031	—

5.1 Assumption that π_{CO} is constant within each release year

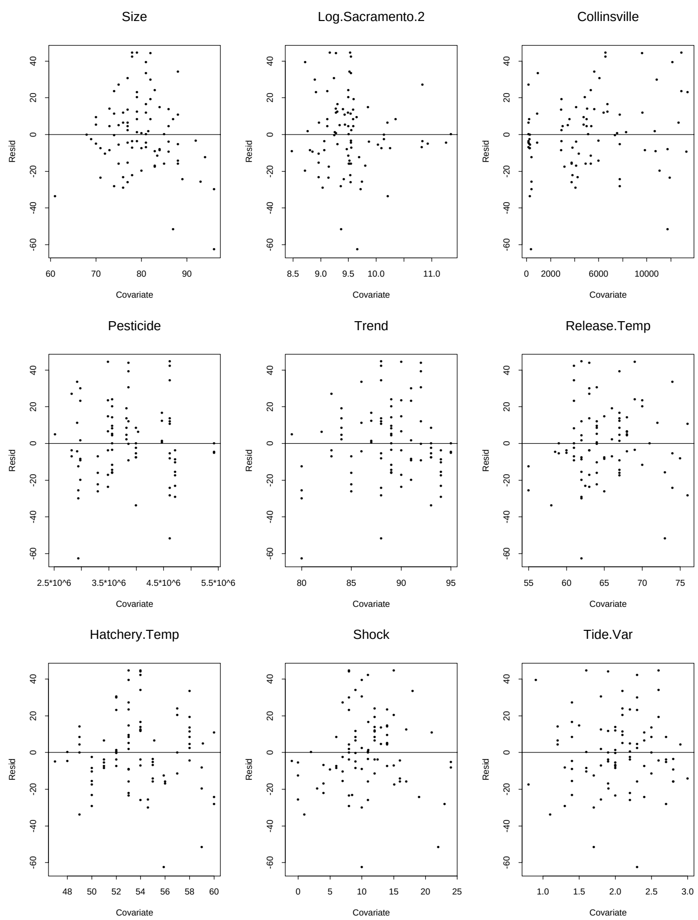
Expanded ocean recoveries of coded wire tagged fish categorized by time, area, and fishery provide information about ocean recovery patterns. We divided these recoveries by the total number of expanded recoveries to form relative recovery patterns. If two release groups have similar migration paths and ocean survival rates, their relative recovery patterns should be similar, consistent with the hypothesis of a constant π_{CO}, whereas very different recovery patterns might cast doubt upon the hypothesis.

Cluster analysis of the relative recovery patterns was carried out on a set of CWT release groups released from the Sacramento system over the period beginning in 1978 and ending in 1992. The resulting dendrograms showed that, in general, similarities within release years were greater than between release years. While this does not prove that the assumption that π_{CO} is constant within each release year, it is consistent with this assumption.

5.2 Within group influence of SI covariates

Table 2 shows the sensitivity of estimates to such correlations by estimating the SI coefficients when one of the SI covariates is omitted. The most notable interactions are between flow and salinity, between pesticides and trend, and amongst the three temperature covariates. Removal of salinity (Collinsville) leads to a sizeable decrease in the flow (Log.Sacramento.2) coefficient, and when flow is removed the salinity coefficient decreases to a lesser degree. When pesticides are removed, the survival trend worsens, but trend's removal has little impact on the pesticide coefficient. The removal of release temperature makes the effect of high hatchery temperatures and large shocks more harmful; conversely, the coefficient for release temperature is little affected by the removal of hatchery temperature or shock.

Figure 4: Residuals for Chipps Island versus SI covariates.



5.3 Residual analyses

The residuals for Chipps Island observations are plotted against the Site Independent covariates in Figure 4 and against the Site Dependent covariates in Figure 5. The residuals versus *Shock* (Figure 4) suggest a possible nonlinearity, in particular a threshold effect, in that for small shocks the model is overestimating the survival. Namely, for shocks less than 7 degrees there are no positive residuals. Both the residuals for releases from Feather River Hatchery were negative, but given such a small sample that is not necessarily alarming; the balance between positive and negative residuals for the remaining release locations seems quite good. No remaining associations seem apparent.

5.4 Influence of individual observations

By leaving one observation out at a time and re-calculating the coefficient estimates, the influence of each observation can be assessed. This was done for 167 observations (excluding 3 observations which were the sole representatives of ocean year effects, 1979, 1981, and

Figure 5: Residuals for Chipps Island versus SD covariates. (For the *.dum* and *.gate* covariates, the value 1.0 means release from that location and gate is open, respectively.)

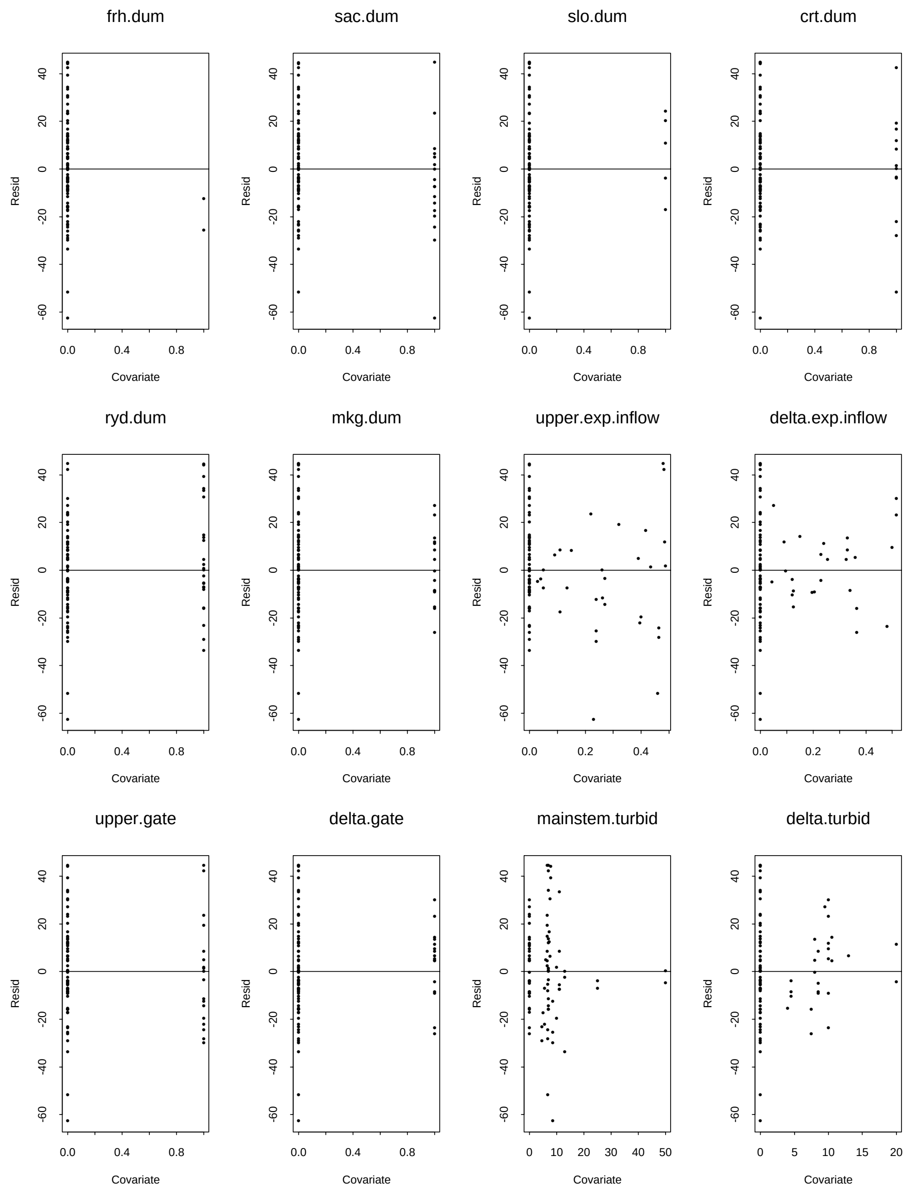
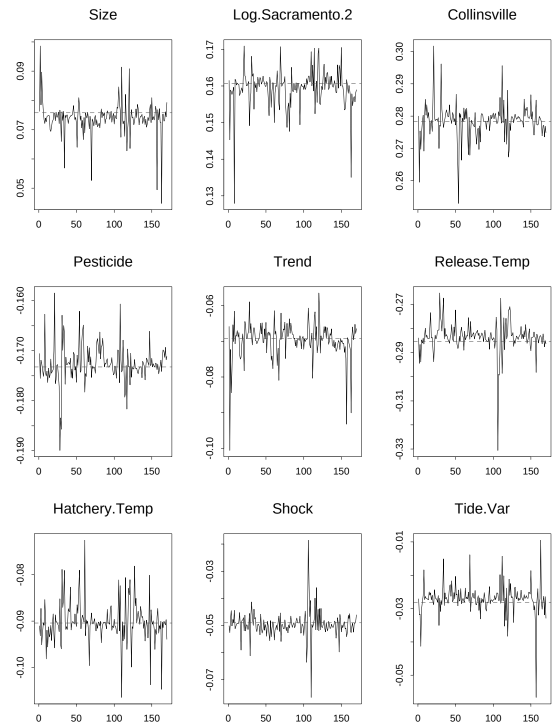


Figure 6: Individual observation's influence on SI coefficients.



1982). Figures 6 and 7 show the estimated coefficients for SI and SD covariates plotted against deleted observation. Note that only in the case *delta.turbid* does the deletion of an observation result in a change in sign.

A summary statistic (similar to Cook's distance measure, (pp. 118–125, Weisberg, 1985)) was calculated as follows:

$$D[i] \quad = \quad \frac{1}{p}\sum_{j=1}^{38}(\hat{\beta}_{\lambda,j,[-i]} - \hat{\beta}_{\lambda,j})^2.$$

The maximum value of D is for the 108th observation, a 1986 ocean recovery observation for a Ryde area release, is 0.00025 (see Figure 8), which indicates a very small change in estimates of $\hat{\beta}$ when this observation is deleted.

5.5 Influence of ocean observations

There is information about covariate effects on smolt mortality present in both recoveries at Chipps Island and in the ocean, although traditionally only the former have been used in modeling. Our analysis tries to gain strength by using both sets of data rather than relying solely on one. In this section we examine the consistency of the two data sets; radical inconsistencies would cause us to doubt the whole enterprise. We also examine the effects of using ocean recoveries in addition to Chipps Island recoveries.

Figure 7: Individual observation's influence on SD coefficients.

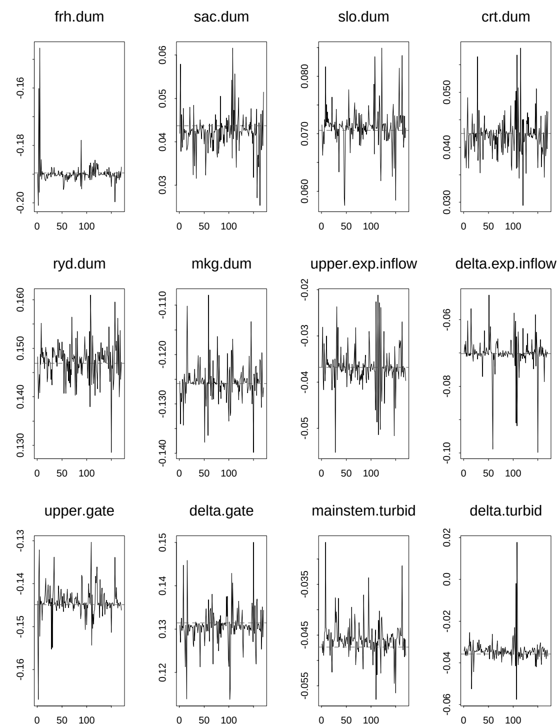


Figure 8: Influence of individual obsn's on coefficients (a variation on Cook's distance).

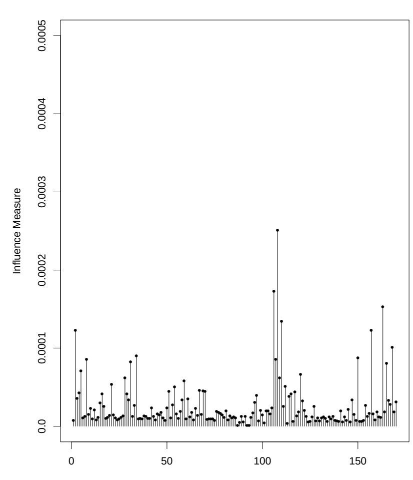


Table 3: Comparing estimates based on Chipps Island observation alone, Ocean observations alone, and Chipps Island and Ocean observations combined.

Covariate	$\hat{\beta}_\lambda$			se($\hat{\beta}_\lambda$)		
	$n_{Chipps,86}$	$n_{Ocean,84}$	$n_{Both,170}$	$n_{Chipps,86}$	$n_{Ocean,84}$	$n_{Both,170}$
Site Independent Factors						
Size	-0.017	0.230	0.076	0.051	0.068	0.051
Log.Sacramento.2	0.109	0.185	0.161	0.055	0.110	0.064
Collinsville	0.310	0.076	0.278	0.054	0.105	0.062
Pesticide	-0.165	-0.263	-0.173	0.049	0.060	0.055
Trend	-0.095	-0.345	-0.069	0.056	0.062	0.065
Release.Temp	-0.215	-0.408	-0.285	0.054	0.091	0.063
Hatchery.Temp	-0.103	-0.111	-0.090	0.057	0.088	0.063
Shock	-0.071	-0.025	-0.049	0.054	0.077	0.060
Tide.Var	0.013	-0.144	-0.028	0.049	0.057	0.046
Site Dependent Factors						
frh.dum	-0.202	-0.145	-0.190	0.057	0.056	0.048
sac.dum	-0.037	0.131	0.044	0.046	0.088	0.052
slo.dum	0.106	-0.022	0.071	0.045	0.075	0.049
crt.dum	0.011	0.039	0.043	0.044	0.082	0.050
ryd.dum	0.164	0.041	0.147	0.046	0.092	0.053
mkg.dum	-0.116	-0.169	-0.126	0.056	0.085	0.060
upper.exp.inflow	-0.068	-0.087	-0.037	0.054	0.095	0.063
delta.exp.inflow	-0.067	-0.066	-0.070	0.058	0.100	0.068
upper.gate	-0.077	-0.173	-0.145	0.056	0.057	0.052
delta.gate	0.089	0.164	0.131	0.057	0.103	0.068
mainstem.turbid	0.013	-0.074	-0.047	0.054	0.101	0.059
delta.turbid	0.010	-0.113	-0.036	0.056	0.102	0.067

Three observation sets are evaluated: $n_{Chipps,86}$, $n_{Ocean,84}$, and $n_{Both,170}$. For $n_{Chipps,86}$ the 86 Chipps Island observations are fit to a 22 parameter model (since the 15 year effect parameters and the Chipps Island indicator can be dropped). For $n_{Ocean,84}$ a 37 parameter model is fit (only Chipps Island indicator is dropped). $n_{Both,170}$ (with 38 parameters) is the original data set used throughout this report. The same ridge parameter value, λ=550, was used in all three cases.

The parameter estimates for commonly used covariates for each of the three sets (along with estimated standard errors) are shown in Table 3. When both the 86 Chipps Island observations and the 84 ocean observations are used simultaneously the estimation program downweights the influence of the ocean data— this partially explains the differences in estimates.

Contrasting $n_{Chipps,86}$ and $n_{Ocean,84}$, the parameter estimates for both sets are plotted against one another in Figure 9. The lower left and upper right quadrants separated by the dotted lines show the estimates that do not change in sign. The estimated coefficients are relatively consistent between the two data sets. Many of the coefficients based on the ocean

Figure 9: Estimated coefficients ($\pm$ 1 se) based on n=84 observations (Ocean data alone) versus n=86 observations (Chipps Island data alone).

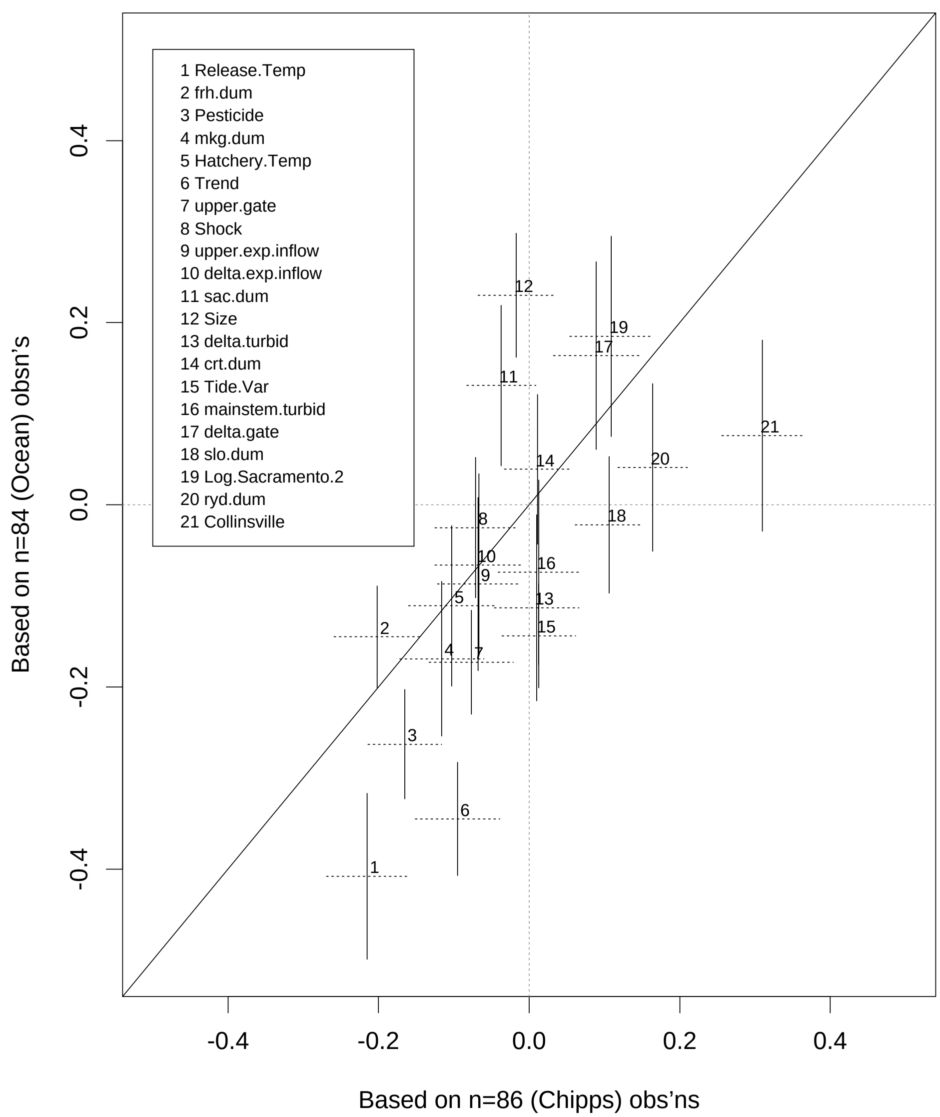
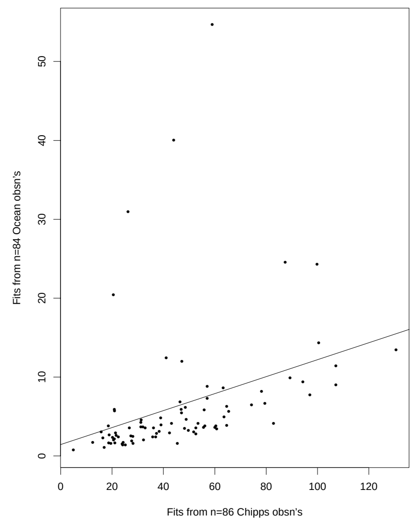


Figure 10: Chipps Island predicted values based on two data sets, n=84 observations (Ocean data alone) versus n=86 observations (Chipps Island data alone). (Straight line across plot is least squares line.)



only data are exaggerations of the coefficients based on Chipps Island data only— negative coefficients become even more negative, and positive coefficients become more positive. At the same time the standard errors for the ocean-only coefficients remain larger, partly a reflection of the larger dispersion parameter. For goodness of fit assessment the fitted Chipps Island recoveries for both models are plotted in Figure 10. Because of different intercepts (affected by different magnitudes of offsets) the fitted values will be at best proportional. With a handful of notable exceptions the fitted values do look roughly proportional— the straight line drawn across the plot is the least squares line. Combining the two data sets for parameter estimation seems appropriate.

Looking at the effects of using ocean recoveries in addition to Chipps Island recoveries now, pairings of the parameter estimates for $n_{Chipps,86}$ are plotted against $n_{Both,170}$ in Figure 11. Again the estimated coefficients are relatively consistent between the two data sets. For goodness of fit assessment fitted values for Chipps Island recoveries based on the two data sets' coefficients are plotted in Figure 12 and the correlation was quite strong. The straight line drawn across the plot is the 45 degree line. Using the average squared Pearson residual as the goodness of fit measure, $n_{Both,170}$ has a slightly worse fit (8.52) compared to $n_{Chipps,86}$ (fit of 7.51). The decrease in quality of fit can be roughly explained as the model based on both data sets is trying to do more, namely account for ocean recoveries, while 'only' adding a 'few' more variables to account for unique ocean effects.

Figure 11: Estimated coefficients (± 1 se) based on n=170 observations (Chipps Island and Ocean data) versus n=86 observations (Chipps Island data alone).

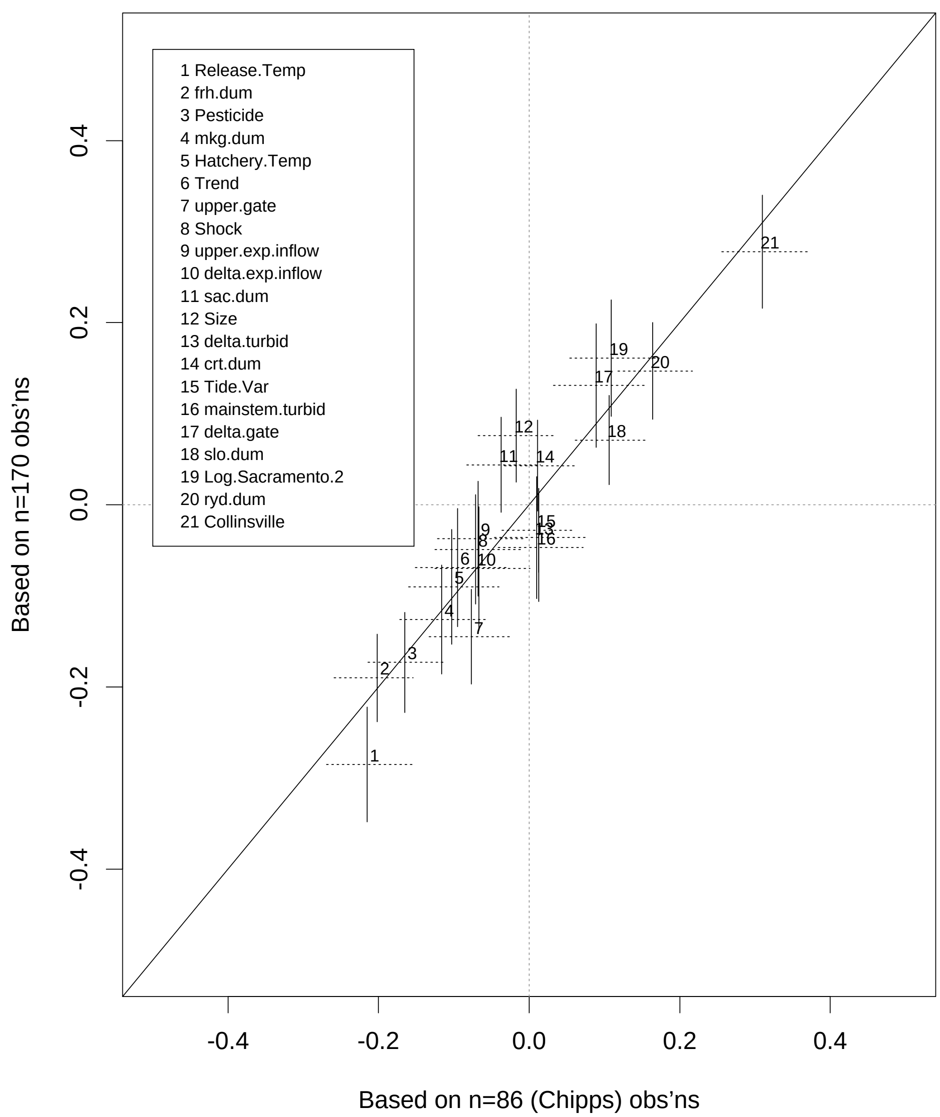
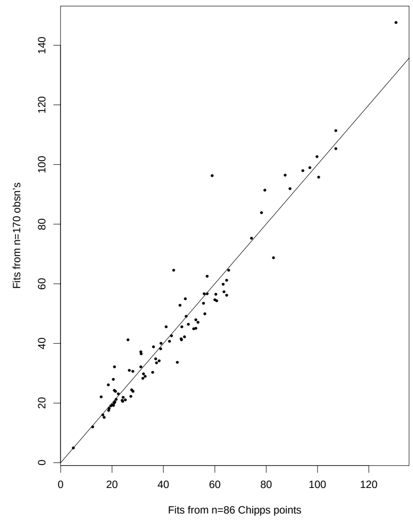


Figure 12: Chipps Island predicted values based on two data sets, n=86 (Chipps Island data) vs n=170 observations (Chipps Island and Ocean data). (Straight line across plot is 45 degree line.)



5.6 Separate size coefficients for Chipps Island and ocean recoveries

One hypothesis suggested by members of the working group was that larger fish may have higher ocean survival rates than smaller fish (note the difference in $Size$ in Table 3 between the Chipps Island data and ocean data results). To examine this hypothesis an additional covariate, an interaction term for Chipps Island indicator variable and fish size, was added. The effect on the other coefficients (excluding $Size$ and the Chipps Island indicator) was miniscule (generally a difference of 0.002 at most). The estimated size coefficient, now for ocean recoveries alone, changed from 0.076 to 0.082 and the coefficient for the new term was -0.070. Thus the size coefficient for Chipps Island recoveries is 0.082 - 0.070 = -0.012. According to these coefficients, larger fish are caught with higher probability in the ocean fisheries than are smaller fish. These results are consistent with the hypothesis, but they are also consistent with a hypothesis that larger fish are more vulnerable to ocean fisheries— if they stayed larger, then they would more often exceed size limits than smaller fish.

5.7 Effect of high flows

The high flow periods were confounded with various factors including gate position and export levels. To examine the influence of observations during high flows, observations with cfs greater than 20,000 cfs were removed. This necessitated the removal of three year indicators, 1982, 1983, and 1993, since releases during these years coincide with high flow periods. Table 4 contrasts the estimated coefficients for the full data set (n=170) and the

Table 4: Coefficients for full data set and with high flow obsn's removed.

Covariate	Full	-Hi Flow
Intercept	-5.982	-6.087
Chipps	1.026	1.040
Site Independent Factors		
Size	0.076	-0.008
Log.Sacramento.2	0.161	0.162
Collinsville	0.278	0.278
Pesticide	-0.173	-0.150
Trend	-0.069	-0.064
Release.Temp	-0.285	-0.268
Hatchery.Temp	-0.090	-0.117
Shock	-0.049	-0.098
Tide.Var	-0.028	-0.037
Site Dependent Factors		
frh.dum	-0.190	-0.250
sac.dum	0.044	-0.068
slo.dum	0.071	0.178
crt.dum	0.043	-0.012
ryd.dum	0.147	0.136
mkg.dum	-0.126	-0.104
upper.exp.inflow	-0.037	-0.032
delta.exp.inflow	-0.070	-0.040
upper.gate	-0.145	-0.132
delta.gate	0.131	0.179
mainstem.turbid	-0.047	0.242
delta.turbid	-0.036	0.051
Year Effects		
1979	-0.103	-0.093
1980	0.087	0.126
1981	-0.059	-0.014
1984	0.127	0.135
1985	0.049	0.054
1986	0.237	0.227
1987	0.243	0.249
1988	0.247	0.270
1989	-0.040	-0.035
1990	-0.035	-0.049
1991	-0.011	-0.001
1992	0.026	0.015

data set with high flows removed (n=145). The most affected coefficient was for mainstem turbidity, the 'low' flow data set implying that increases in turbidity increased recovery rates. Lesser changes were for *sac.dum* and *slo.dum*, with releases in the Sacramento group predicted to do more poorly and releases in the Sutter and Steamboat Sloughs doing better (relative to the baseline site, Jersey Point).

5.8 Effect of alternative export measure

Some managers of the Sacramento river system prefer to evaluate water exports in absolute values rather than as a ratio of system flow. Because 'in'-flow is in the model (the *Log.Sacramento.2* covariate) using exports/inflow ratio or total exports should yield nearly equivalent results. This can be seen in Table 5. The effects on other coefficients and on goodness of fit were minor, with exports/inflow ratio giving a slightly better fit to the Chipps Island observations (average squared Pearson error of 8.72 versus 8.52).

Table 5: Coefficients for models using export/inflow ratio and using absolute export volume.

Covariate	Ratio	Absolute
Intercept	-5.982	-5.983
Chipps	1.026	1.028
Site Independent Factors		
Size	0.076	0.083
Log.Sacramento.2	0.161	0.172
Collinsville	0.278	0.269
Pesticide	-0.173	-0.168
Trend	-0.069	-0.071
Release.Temp	-0.285	-0.283
Hatchery.Temp	-0.090	-0.092
Shock	-0.049	-0.051
Tide.Var	-0.028	-0.030
Site Dependent Factors		
frh.dum	-0.190	-0.186
sac.dum	0.044	0.055
slo.dum	0.071	0.077
crt.dum	0.043	0.049
ryd.dum	0.147	0.162
mkg.dum	-0.126	-0.134
upper.exp.inflow	-0.037	-0.028
delta.exp.inflow	-0.070	-0.020
upper.gate	-0.145	-0.146
delta.gate	0.131	0.109
mainstem.turbid	-0.047	-0.052
delta.turbid	-0.036	-0.033
Year Effects		
1979	-0.103	-0.102
1980	0.087	0.087
1981	-0.059	-0.059
1982	0.062	0.063
1983	-0.069	-0.068
1984	0.127	0.128
1985	0.049	0.048
1986	0.237	0.241
1987	0.243	0.244
1988	0.247	0.248
1989	-0.040	-0.038
1990	-0.035	-0.034
1991	-0.011	-0.014
1992	0.026	0.031
1993	0.239	0.240

6 Discussion

6.1 Caveats

Before further interpreting the results, we give three warnings:

1. The coefficients cannot be viewed in strict isolation. For example, we cannot say that a unit increase in (standardized) log Sacramento flow leads to an increase of 0.161 in the (log) survival/recovery rate. This is partially due to correlation between some of the covariates. Sacramento flow and salinity at Collinsville have a strong inverse relationship, for instance, as flow increases salinity tends to decrease. Therefore a unit increase in flow might mean a certain amount of decrease in salinity which in turn would affect the survival rate. Section 5.2 addressed this issue further.

   The coefficient of an individual variable depends upon which variables are in and not in the model. Table 2 showed examples of the relationships between variables in the the model. On the other hand, a variable in the model can be a proxy[7] for an unmeasured variable.

2. Predictive ability outside the range of data observed is completely unknown, i.e., it is dangerous to extrapolate beyond the data base. For example, gates have never been left open when Sacramento flows exceed 20,000 cfs[8], so predicted survival for open gates and flows of 30,000 cfs could be very misleading.

3. Predictive ability for the survival of future releases depends upon other factors not included in the model staying similar. Important conditions that have held historically need to remain in the future else the model may not be applicable.

To help determine the allowable range of the data refer to the plots in Appendix A showing combinations of covariates. For example, the plot of $Collinsville$ versus $Log.Sacramento.2$ (Figure 17) clearly shows the correlation and highlights the fact that a very high Sacramento flow would not have a very high salinity value. An alternative approach to determining whether or not a proposed set of covariates falls within the range of data is to determine how close the proposed set is to the historical set[9].

[7]A proxy variable is a variable 'strongly' associated with an unmeasured variable not included in the model. The unmeasured variable is what is actually causing an observed effect.

[8]The 20,000 cfs ceiling is particular to the data used for model development. The ceiling value including data not used in the model is 25,000 cfs.

[9]First, the Euclidean distances between each of the 86 Chipps Island observation covariate vectors is calculated. Namely, for observations j and k, on a covariate by covariate basis subtract the standardized value for observation j from the standardized value for observation i, square each distance, sum the squares, and then take the square root of the sum— that gives the distance between the 2 'points' in the covariate space. Second, the distance to each observation's nearest neighbor is selected, which yields a distribution of nearest neighbor distances. Third, find the nearest neighbor for the proposed covariate vector (from the 86 Chipps Island observations) and compare the resulting distance to to the distribution of nearest neighbor distances. If its nearest neighbor distance falls within the middle 95% of the distribution, say, the proposed covariate combination might be thought reasonable.

6.2 Conclusions

1. Each of the three collections of covariates, SI, SD, and Year effect, are all individually important at a group level in predicting survival through the delta. Year effect should be viewed as something necessary in any model, controlling for known variations in ocean environmental conditions and fishing regimes. If SI or SD are dropped from the model, the quality of the model fit to Chipps Island recoveries decreases drastically (both in practical terms and statistically). The table below gives the average squared Pearson residual for the 86 Chipps Island recoveries, $\sum_{i=1}^{86} \frac{(Y_{C,i}-\hat{Y}_{C,i})^2}{\hat{Y}_{C,i}}$, as well as the degrees of freedom for error[10] for different models:

Covariates	Avg. Error	Error *df*	Avg Err/*df*
Year	24.4	153	0.16
Year+SI	15.1	144	0.11
Year+SD	17.4	141	0.12
Year+SI+SD	8.5	132	0.06

The ratio of error to error degrees of freedom is somewhat analogous to the mean square error in standard ANOVA. Including SI or SD results in a 25 to 31% reduction in average error, while including both SI and SD yields a 62% reduction.

2. The SI components can be cautiously interpreted, on the basis of their magnitude on the standardized scale and standard errors (see Figure 3), as follows:

As release temperature and hatchery temperature increase over the ranges observed (see Appendix section A.4), fish survival decreases. The two temperatures are highly correlated since hatchery water and river water usually come from the same source; so it is simplest to say that as the river water warms, mortality increases.

Increased levels of (rice) pesticides are associated with increased mortality. Pesticides are yearly values and as such may be proxies for some unmeasured variable, although the inclusion of the pesticide covariate did account for variation over and above that captured by an annual linear trend term.

Flow and salinity are important 'positive' factors. Because the two are inversely related, when flow is high, salinity is low, they should not be viewed in isolation— but their combined effect is significant.

The impacts of fish size, the tidal factor, and temperature shock are slight given the other covariates' presence in the model.

The most successful and least successful release groups, in terms of fitted rates, $S_C q$, were compared in terms of their SI covariate values. Figure 13 contains boxplots of the values for eight SI covariates of the ten 'best' and ten 'worst' releases and the eight SI covariates. The points are labelled by their relative recovery rates, with 1 being best and 86 being worst. There is a fair amount of overlap in covariate values for the two

[10] The degrees of freedom for error are overestimates given that a ridge parameter was used, but the overall conclusion should not be affected.

Table 6: Covariate values for the 10 best (listed first) and 10 worst release groups based on model estimates of recovery 'rate' at Chipps Island.

Recovery 'Rate'	Fish Size	(log) Sac Flow	Collin. Salin.	Pest.	Trend	Rel. Temp.	Hatch. Temp.	Shock	Tide Var	Site	Export	Gate Pos.	Turbidity
1.87	86	8.85	13301	2982796	91	61	53	5	2.8	Jers	0.20	Open	8.5
1.50	81	10.82	166	2821676	83	61	55	4	2.0	Ryde	—	—	25.0
1.27	79	10.84	164	2821676	83	60	52	8	2.7	Crt	0.04	Closed	25.0
1.17	82	9.28	9611	3862071	92	63	54	8	1.9	Ryde	—	—	7.8
1.16	82	8.89	10846	2982796	91	63	52	8	2.3	Jers	0.52	Open	10.0
1.13	86	9.87	4127	3484814	90	63	54	12	1.9	Slou	—	—	—
1.13	86	9.12	4999	3559856	89	62	49	8	2.9	Ryde	—	—	6.5
1.11	79	8.92	13404	2982796	91	61	52	8	2.3	Mk-G	0.52	Open	10.0
1.09	83	9.81	4127	3484814	90	63	56	4	1.9	Slou	0.00	—	—
1.07	79	9.16	12873	3484814	90	69	53	15	1.6	Ryde	0.00	—	6.5
0.13	74	9.36	7752	4613002	88	76	60	23	2.7	Crt	0.46	Open	6.7
0.19	89	9.42	7752	4613002	88	74	60	19	2.4	Sac	0.46	Open	6.7
0.25	74	9.54	891	2921654	86	72	58	12	2.1	Mk-G	0.24	Open	20.0
0.27	73	9.07	2879	4713055	94	62	50	9	2.0	Mk-G	0.12	Closed	4.5
0.30	94	9.69	422	2940220	80	55	56	0	1.7	FRH	0.24	Open	8.5
0.30	85	9.50	5113	3559856	89	71	57	11	2.5	Crt	0.26	Open	7.0
0.31	77	9.55	899	2921654	86	68	58	16	2.6	Mk-G	0.23	Open	20.0
0.31	77	8.96	3745	4713055	94	62	50	7	1.4	Mk-G	0.12	Closed	4.0
0.32	83	9.50	5334	3559856	89	70	57	12	2.5	Sac	0.26	Open	7.0
0.33	88	9.43	7752	4613002	88	76	60	21	2.4	Slo	—	—	—

sets of extreme cases, but overall differences are apparent for those covariates that had relatively large coefficients. Note that the impossibility of separating salinity from flow is apparent from the relatively high salinity values combined with low flows for the ten best. Table 6 shows several of the covariate values (including SD covariates) for these twenty groups. Seeing the combination of covariates for a given release provides more clues as to why a group did well or poorly. For example, the most successful group, from Jersey Point in 1991, had low flow, but high salinity (and note that the gates were open for this central delta release). Table 6 also suggests that site effects were quite influential in that four of the best groups came from Ryde (and *ryd.dum* had the largest positive site indicator coefficient).

The above ranking was repeated after partially removing the site effects by ranking after dividing by the fitted recovery rates by $\exp(\hat{\beta}_{11} frh.dum + \ldots + \hat{\beta}_{16} mkg.dum$ (see Table 7 and Figure 14). Because of interactions between other site dependent covariates (export fractions, gate position, and turbidity) this attempt to control for site effect is only partially effective but nonetheless revealing. In Table 7 are also given the contributions per covariate, namely $\hat{\beta}_i x_i$, to the estimated fit. The negative effect of being released from Feather River Hatchery and the positive effect of being released from Ryde are large enough that when these site effects are 'controlled for', releases formerly seen as 'successful' or 'unsuccessful' on the raw scale become relatively unsuccessful or successful. The most noticeable separation is from the *Shock* covariate and relatedly from *Release.Temp* and *Hatchery.Temp*. Somewhat higher flows, higher salinity, and lower pesticide levels are evident for the best compared to the worst.

3. The SD components require more careful examination because of the way interactions

Figure 13: Dot plots of SI covariates for the 10 best and 10 worst releases, i.e., releases with highest and lowest estimated Chipps Island recovery rates.

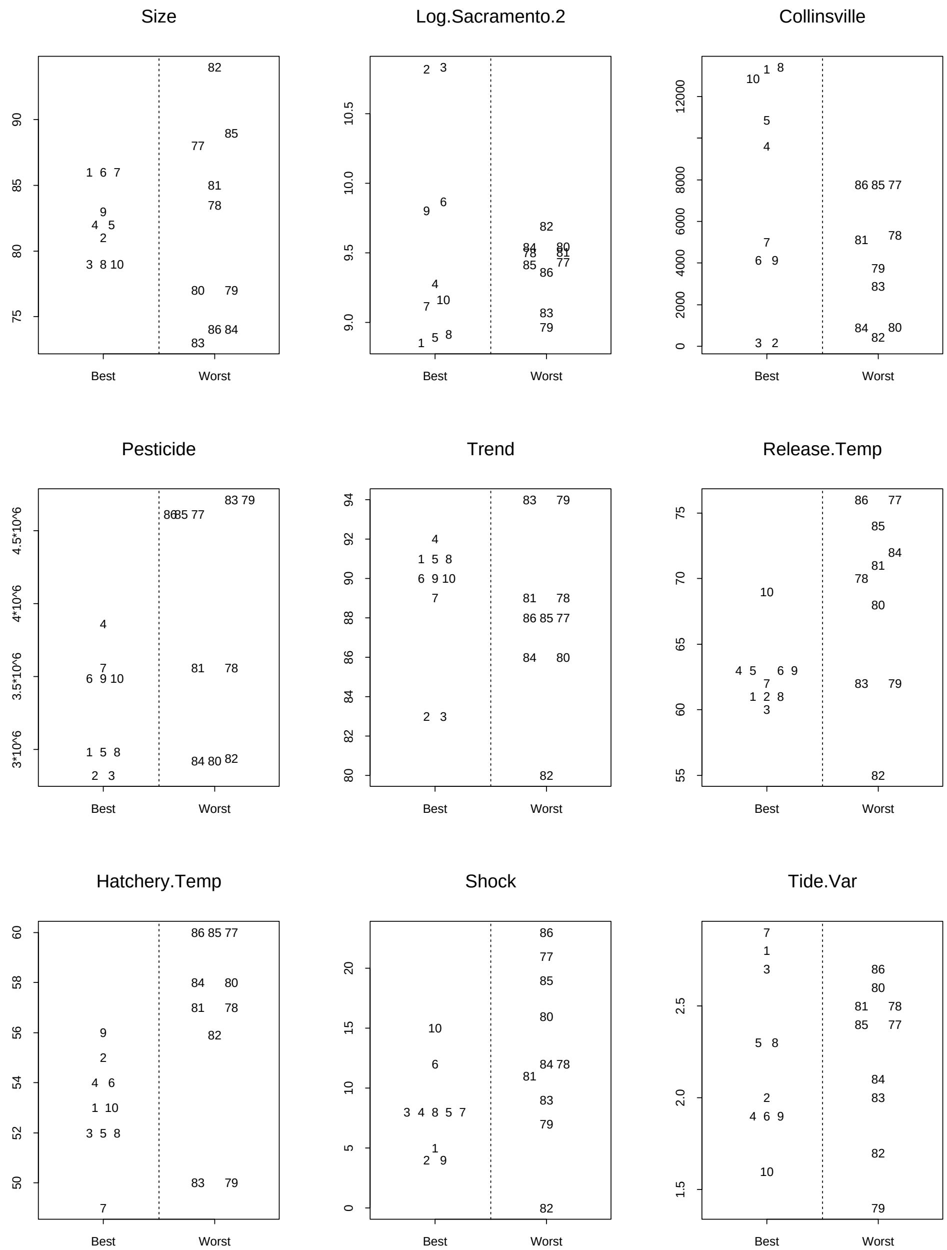
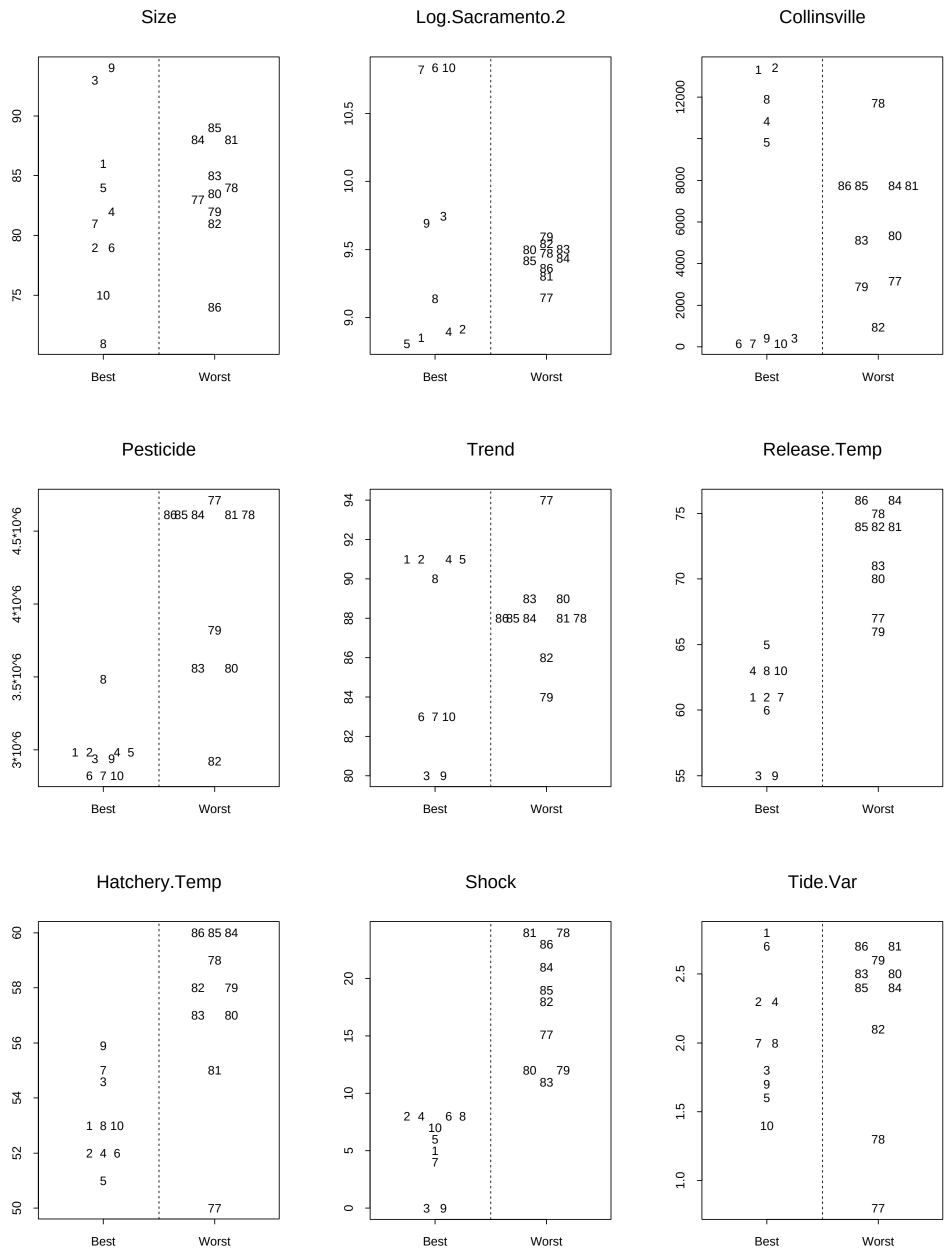


Table 7: Covariate values for the 10 best (listed first) and 10 worst release groups based on model estimates of recovery 'rate' at Chipps Island with site effect removed. Numbers in smaller, italicized type are the contribution to each estimate, the coefficient times covariate value.

Recovery 'Rate'	Fish Size	(log) Sac Flow	Collin. Salin.	Pest.	Trend	Rel. Temp.	Hatch. Temp.	Shock	Tide Var	Site	Export	Gate Pos.	Turbidity
1.87	86	8.85	13301	2982796	91	61	53	5	2.8	Jers	0.20	Open	8.5
	(*1.01*	*2.79*	*0.96*	*-0.80*	*-1.63*	*-3.87*	*-1.55*	*-0.05*	*-0.16*		*-0.10*	*0.34*	*-0.06*)
1.54	79	8.92	13404	2982796	91	61	52	8	2.3	Mk-G	0.52	Open	10.0
	(*0.93*	*2.81*	*0.97*	*-0.80*	*-1.63*	*-3.87*	*-1.52*	*-0.08*	*-0.13*		*-0.25*	*0.34*	*-0.07*)
1.52	93	9.75	422	2940220	80	55	55	0	1.8	FRH	0.24	Closed	8.5
	(*1.09*	*3.07*	*0.03*	*-0.79*	*-1.43*	*-3.49*	*-1.59*	*0.00*	*-0.10*		*-0.05*	*0.00*	*-0.06*)
1.16	82	8.89	10846	2982796	91	63	52	8	2.3	Jers	0.52	Open	10.0
	(*0.96*	*2.81*	*0.79*	*-0.80*	*-1.63*	*-3.99*	*-1.52*	*-0.08*	*-0.13*		*-0.25*	*0.34*	*-0.07*)
1.14	84	8.81	9843	2982796	91	65	51	6	1.6	Mk-G	0.34	Open	8.5
	(*0.98*	*2.78*	*0.71*	*-0.80*	*-1.63*	*-4.12*	*-1.49*	*-0.06*	*-0.09*		*-0.17*	*0.34*	*-0.06*)
1.13	79	10.84	164	2821676	83	60	52	8	2.7	Crt	0.04	Closed	25.0
	(*0.93*	*3.42*	*0.01*	*-0.75*	*-1.48*	*-3.80*	*-1.52*	*-0.08*	*-0.15*		*-0.01*	*0.00*	*-0.17*)
1.08	81	10.82	166	2821676	83	61	55	4	2.0	Ryde	—	—	25.0
	(*0.95*	*3.41*	*0.01*	*-0.75*	*-1.48*	*-3.87*	*-1.60*	*-0.04*	*-0.11*		—	—	*-0.17*)
1.07	71	9.14	11897	3484814	90	63	53	8	2.0	Jers	0.48	Open	10.0
	(*0.83*	*2.88*	*0.86*	*-0.93*	*-1.61*	*-3.99*	*-1.55*	*-0.08*	*-0.11*		*-0.24*	*0.34*	*-0.07*)
1.03	94	9.69	422	2940220	80	55	56	0	1.7	FRH	0.24	Open	8.5
	(*1.10*	*3.06*	*0.03*	*-0.79*	*-1.43*	*-3.49*	*-1.63*	*0.00*	*-0.10*		*-0.05*	*-0.35*	*-0.06*)
1.02	75	10.84	168	2821676	83	63	53	7	1.4	Mk-G	0.05	Closed	9.5
	(*0.88*	*3.42*	*0.01*	*-0.75*	*-1.48*	*-3.99*	*-1.55*	*-0.07*	*-0.08*		*-0.02*	*0.00*	*-0.07*)
0.12	74	9.36	7752	4613002	88	76	60	23	2.7	Crt	0.46	Open	6.7
	(*0.87*	*2.95*	*0.56*	*-1.23*	*-1.57*	*-4.82*	*-1.75*	*-0.22*	*-0.15*		*-0.10*	*-0.35*	*-0.05*)
0.17	89	9.42	7752	4613002	88	74	60	19	2.4	Sac	0.46	Open	6.7
	(*1.04*	*2.97*	*0.56*	*-1.23*	*-1.57*	*-4.69*	*-1.75*	*-0.18*	*-0.14*		*-0.10*	*-0.35*	*-0.05*)
0.24	88	9.43	7752	4613002	88	76	60	21	2.4	Slo	—	—	—
	(*1.03*	*2.98*	*0.56*	*-1.23*	*-1.57*	*-4.82*	*-1.75*	*-0.20*	*-0.14*		—	—	—)
0.27	85	9.50	5113	3559856	89	71	57	11	2.5	Crt	0.26	Open	7.0
	(*1.00*	*3.00*	*0.37*	*-0.95*	*-1.59*	*-4.50*	*-1.66*	*-0.10*	*-0.14*		*-0.06*	*-0.35*	*-0.05*)
0.27	81	9.54	951	2921654	86	74	58	18	2.1	Ryde	—	—	11.0
	(*0.95*	*3.01*	*0.07*	*-0.78*	*-1.54*	*-4.69*	*-1.69*	*-0.17*	*-0.12*		—	—	*-0.07*)
0.28	88	9.31	7752	4613002	88	74	55	24	2.7	Ryde	—	—	6.7
	(*1.03*	*2.94*	*0.56*	*-1.23*	*-1.57*	*-4.69*	*-1.60*	*-0.23*	*-0.15*		—	—	*-0.05*)
0.28	83	9.50	5334	3559856	89	70	57	12	2.5	Sac	0.26	Open	7.0
	(*0.98*	*3.00*	*0.39*	*-0.95*	*-1.59*	*-4.44*	*-1.66*	*-0.11*	*-0.14*		*-0.06*	*-0.35*	*-0.05*)
0.30	82	9.60	2902	3820103	84	66	58	12	2.6	Crt	0.32	Open	6.5
	(*0.96*	*3.03*	*0.21*	*-1.02*	*-1.50*	*-4.18*	*-1.69*	*-0.11*	*-0.15*		*-0.07*	*-0.35*	*-0.04*)
0.34	84	9.47	11721	4613002	88	75	59	24	1.3	Ryde	—	—	6.7
	(*0.98*	*2.99*	*0.85*	*-1.23*	*-1.57*	*-4.76*	*-1.72*	*-0.23*	*-0.07*		—	—	*-0.05*)
0.34	83	9.15	3161	4713055	94	67	50	15	0.8	Sac	0.11	Closed	5.0
	(*0.97*	*2.89*	*0.23*	*-1.26*	*-1.68*	*-4.25*	*-1.46*	*-0.14*	*-0.05*		*-0.02*	*0.00*	*-0.03*)

Figure 14: Dot plots of SI covariates for the 10 best and 10 worst releases, i.e., releases with highest and lowest estimated Chipps Island recovery rates (site effects removed).



have been built into the model. The simplest thing to do for particular combinations of SD factors is to calculate ratios of survival rates as was done in the example (22), as well as calculate standard errors of the ratios.

Figure 3 and, in the case of dummy variables thinking about default values, provides some guidance as to importance of various covariates. Beginning with release location, relative to Jersey Point (the default) releases from Ryde clearly do better, while releases far upstream at Feather River Hatchery, or higher up in the central delta in the Mokelumne-Georgiania cluster, do poorly. Poorer survival from the Mokelumne-Georgiana area would be consistent with a hypothesis that fish released closer to the western delta tend to get 'lost' more often, not finding the mainstem, than do fish released lower in the central delta nearer the mainstem re-entry point. The other "mainstem" release points, Sacramento, Courtland, and the Sutter-Steamboat Slough combination, are roughly equivalent with slight evidence toward being better than Jersey Point. In summary the site effects are for the most part relatively weak, but in composite they tell a consistent and plausible story.

Gate position seems influential for releases upstream of the cross-channel gate (*upper.gate* coefficient), with the gate being open increasing mortality, as some outmigrating smolts on the mainstem Sacramento river are presumably diverted into the central delta. Conversely for fish released in the central delta (*delta.gate*) having the gates open appears to have some positive effect— perhaps the additional water entering the central delta increases outmigration speed? Referring to Table 6, of the four best groups affected by gate position, three were central delta groups with the gate open and the other (Courtland) had the gate closed. For the five mainstem releases in the worst group, the gates were always open. Unfortunately, the absence of data for situations with high flows and an open gate limits the strength of the conclusions.

The effect of exports remains quite ambiguous— the coefficients *upper.exp.inflow* and *delta.exp.inflow* are slightly negative, but the approximate 95% confidence intervals have considerable overlap with positive values.

7 References

Britton, A. (1996.) (unpublished memorandum) 'Updated Salmon Smolt File'.

Kjelson, M., Greene, S., and Brandes, P. (1989.) "A model for estimating mortality and survival of fall-run chinook salmon smolts in the Sacramento river delta between Sacramento and Chipps Island." USFWS Technical Report.

McCullagh, P., and J. A. Nelder. (1989.) *Generalized linear models, 2nd Ed.*. Chapman and Hall. London.

Weisberg, S. (1985.) *Applied Linear Regression, 2nd Ed.* Wiley and Sons, New York.

A Data details

A.1 Overview

The basic observations are numbers of recoveries by a trawl sweeping an area near Chipps Island and numbers of estimated recoveries in the ocean fisheries (at ages 2, 3, and 4). For most releases then there are two observations, recoveries at Chipps Island and estimated ocean recoveries; e.g., for a 1979 Sacramento release 50 fish were caught at Chipps Island and 98 were estimated to be caught in the ocean fisheries. Ocean recoveries come from a large number of samples taken from catches landed all along the Pacific coast over a three year period. Since catches are sampled at varying rates, we use the total number of expected recoveries—observed recoveries divided by sampling fraction— rather than observed recoveries.

The original data set (provided by Britton (1996)) contains recovery information for 152 releases. The data set was reduced by deleting the Port Chicago releases, one Coleman National Fish hatchery group released at Discovery Park and all Old River releases. The Port Chicago releases generally only contribute to the ocean fishery since the release point is downstream of Chipps Island (with additional assumptions they could be utilized to provide a means of separating survival rates from recapture probabilities). The Coleman group was deleted because of its exceptional nature and the Old River releases were deleted because their extreme location in the central delta (quite near export pumps) would require treatment quite different than other central delta releases and introduce even more parameters.

Many releases, differing essentially in tag code alone, were collapsed into single groups with releases and recoveries appropriately aggregated.

An initial fitting of the model to the data revealed one clear outlier, a single Chipps Island recovery from a collapsed pair of releases made in 1981 from Discovery Park. After deletions and aggregations, there were a total of n=170 observations, 86 observations of Chipps Island recoveries and 84 observations of ocean recoveries.

A.2 Variable definition

Variable definitions largely follow Britton (1996), who prepared the original data files. Sources are indicated by italicized print. DPR is Department of Pesticide Regulation and DWR is Department of Water Resources.

1. Y_C (*USFWS*): number of recoveries in Chipps Island trawl fishery;
2. Y_O (*USFWS*): ocean recoveries in years 2–4 following release adjusted by sampling level;
3. R (*USFWS*): number of tagged fish released;

4. Release Site (*USFWS*): location of release, all the site names in the original data file are listed here,

`Feather River Hatchery`	`Port Chicago`	`Discovery Park`
`Rio Vista`	`Discovery Park (FRH)`	`Discovery Park (CNHF)`
`Courtland`	`Lower Old River`	`Lower Mokelumne`
`Isleton (Ryde)`	`South Fork Mokelumne`	`Ryde`
`North Fork Mokelumne`	`Miller Park`	`Benicia`
`Steamboat Slough`	`Jersey Point`	`Sutter Slough`
`Courtland (east)`	`Courtland (west)`	`Georgiana Slough`

5. f_C (*USFWS*): the fraction of the estimated length of outmigration time and area swept by the trawl at Chipps Island;

6. f_O: the ratio of estimated total recoveries in ocean catches to observed total recoveries in samples of catches;

7. *Size* (*USFWS*): an indirect measurement of average fish size determined by water displacement and converted to mm (or occasionally based on subsamples of length);

8. *log.Sacramento.*2 (*Phyllis Fox, DWR data*): the natural logarithm of the median flow at Freeport in cfs from release date to last day of recoveries at Chipps Island;

9. *Collinsville* (*Trihey & Associates*): average conductivity at Collinsville, measured in micro mho/cm [11], at Collinsville for the period from two days before the first day of recovery to the last day of recoveries at Chipps Island;

10. *Pesticide* (*DPR data*): annual amount of applied rice pesticide in pounds from Pesticide Use Report, Indexed by Commodity, 1979-1995;

11. *Trend*: (last 2 digits of) year of release;

12. *Release.Temp* (*USFWS*): river temperature in degrees Fahrenheit taken at time of release, near the shore and at the surface;

13. *Hatchery.Temp* (*Phyllis Fox, DWR data*): mean temperature in degrees Fahrenheit at Feather River Hatchery on release date;

14. *Shock* (*USFWS*): Truck temperature - Release temperature, where truck temperature, in degrees Fahrenheit, is the water temperature in the truck carrying smolts and is measured at time of release;

[11] A mho is the inverse of an ohm. A mho is thus a measure of electrical conductivity and the higher the salinity, the higher the conductivity (Tom Taylor, personal communication.)

15. *Tide.Var* (*Buell & Associates*): sum of the early tidal 'asymmetry' factor and the early tidal 'trend' factor, where assymetry is the absolute difference between the low-low and high-low tides on the day midway between the release date and the first day of recoveries at Chipps Island and trend is the difference between the tidal asymmetry factor for low tides on the day before and the day after the date of the early tidal asymmetry factor;

16. $---.exp.inflow$ (*Phyllis Fox, DWR data*): median daily values for period from release date to last day of recovery at Chipps Island for the ratio of State Water Project (SWP) and the Central Valley Project (CVP) to the Sacramento flow, Yolo Bypass flow, San Joaquin flow and miscellaneous East Side Stream flow

17. $---.gate$ (*Phyllis Fox, DWR data*): Delta cross channel gate position— coded as 1 for open and 0 for closed;

18. *mainstem.turbid* (*USFWS*): average FTUs (Formazine Turbidity Unit), calculated at Greene's Landing (near Courtland) for the period from release date to last day of recoveries at Chipps Island;

19. *delta.turbid* (*USFWS*): average FTUs, calculated at Potato Point for the period from release date to last day of recoveries at Chipps Island.

A.3 Combining release groups

In combining release groups, covariate values sometimes differed slightly and average covariate values weighted by release numbers were used. For example, there were two replicate groups released in 1991 from Discovery Park (Codes are 6-1-14-2-7 and 6-1-14-2-8) numbering 51,392 and 51,272. The average sizes for the two groups were 80 and 83 mm; the weighted average is 81.50.

The one exception to the weighted averaging was Chipps Island fishing effort. Instead the maximum value within a replicate set was chosen. Assuming replicate groups are indeed replicates, the fact that one group is recovered a day earlier than another group, say, should be due to chance alone. Using the maximum space-time percentage is an attempt to more accurately measure the fishing pressure experienced by the entire set of replicates.

The aggregated release groups are listed below by release year and location:

Year	Location	Tag Codes
1981	Discovery Park	6-62-14, 6-62-17
1984	Ryde	6-42-9, 6-62-29
1985	Courtland	6-62-38 6-62-39 6-62-40 6-62-41
1987	Courtland	6-62-56,6-62-57
1987	Courtland	6-62-53 6-62-54
1988	Steamboat Slough	6-31-5 6-31-6
1988	Miller Park	6-62-61 6-62-62
1988	Courtland	6-62-59 6-62-60
1988	Courtland	B6-14-4 B6-14-5
1988	Miller Park	B6-14-6 B6-14-7
1988	Courtland	B6-14-2 B6-14-3
1989	Courtland (west + east)	6-58-5 6-1-14-1-3
1989	Miller Park	6-31-15 6-31-17
1991	Miller Park	6-1-14-2-9 6-31-24
1991	Miller Park	6-1-14-2-7 6-1-14-2-8

A.4 Data summaries

The following summary statistics are based on values for the 86 Chipps Island observations. Exceptions are Oceans.recs and Ocean.expansion, based on 84 observations, because more recent observations are not available.

Variable	Min	1st Quartile	Median	Mean	3rd Quartile	Maximum
Chipps Isl. Recoveries	3.0	23.0	36.5	46.0	63.2	159.0
Ocean Recoveries	10.0	110.8	272.0	437.3	573.8	1983.0
Number Released	14,920	50,620	52,680	65,270	94,130	160,200
Chipps Isl. Effort	0.07	0.10	0.11	0.12	0.11	0.25
Ocean Expansion	2.9	4.0	4.4	4.5	4.9	6.1
Size	61.0	76.0	79.0	79.9	83.9	96.0
Sacramento Flow	4805	10,310	13,350	16,670	15,300	84,750
Collinsville	164	2875	4797	5131	7477	13,400
Pesticide (1000s lbs)	2516	3485	3820	3826	4471	5422
Release Temp	55.0	62.0	64.0	65.1	67.8	76.0
Hatchery Temp	47.0	52.0	53.5	53.7	55.9	60.0
Truck Temp	48.0	52.0	54.5	54.4	56.0	60.0
Shock	-1.0	8.0	11.0	10.7	14.0	24.0
Tide Variable	0.80	1.70	2.00	2.02	2.38	3.00
Export/Inflow Ratio	0.024	0.12	0.24	0.26	0.40	0.52
Exports	1276	2473	4502	4415	5723	10,180
Mainstem Turbidity	4.0	6.5	7.0	9.2	8.5	50.0
Central Turbidity	4.0	8.5	10.0	11.2	13.0	20.0

Figure (15) plots adjusted Chipps Island recovery rates against all the covariates. Adjusted recovery rate is the number of Chipps Island recoveries divided by release number and the estimated percentage of space-time sampled by the trawl.

Similarly Figure (16) plots adjusted ocean recovery rates against all the covariates. Adjusted ocean recovery rates are estimated number of age 2, 3, and 4 recoveries in ocean fisheries divided by number released and the average catch sample expansion factor.

Figures (17) and (18) contains plots of the site independent covariates against one another. Figure (19) shows plots of the site dependent covariates against one another. Lastly, Figures (20) and (21) are plots of the site independent covariates against the site dependent covariates.

A.5 Data retrieval

S-PLUS data objects labelled `Sac.chinook` and `model.mat` can be downloaded at
http://www.uidaho.edu/˜newman/usfwsglm.html
`Sac.chinook` is a data frame of 116 release groups (after collapsing over replicates) and 20 variables with Chipps Island and ocean recoveries as two separate variables. Some rows in `Sac.chinook` are missing values. `model.mat` is a data frame with 170 observations, 86 for Chipps Island recoveries, and 84 for ocean fishery recoveries. `model.mat` was the data object used for model fitting and contains dummy variables and other created variables, e.g., $log.Sacramento.2$.

Figure 15: (Effort Adjusted) Chipps Island Recovery Rate vs Covariates. (Pesticide in standard units.)

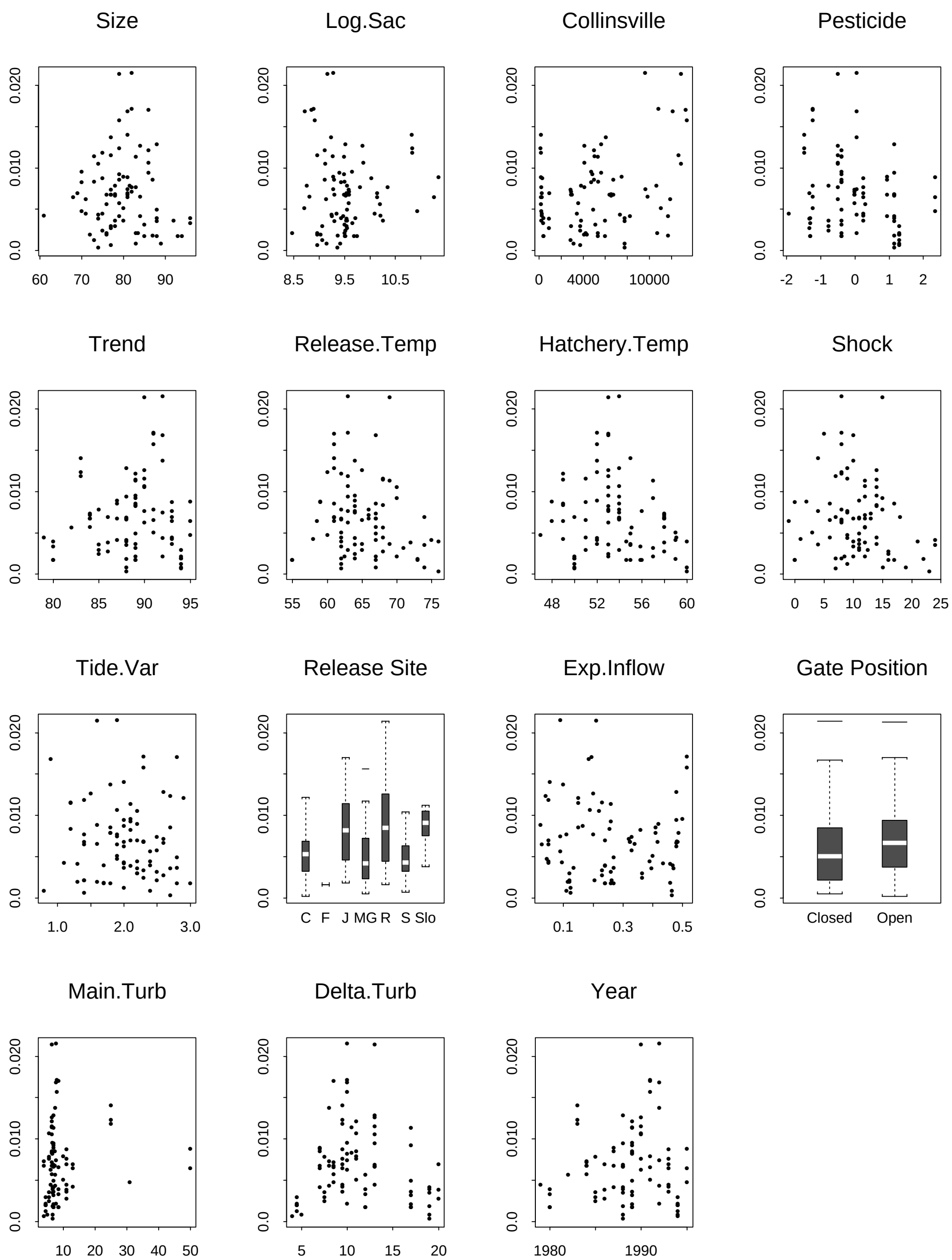


Figure 16: (Effort Adjusted) Ocean Recovery Rate vs Covariates. (Pesticide in standard units.)

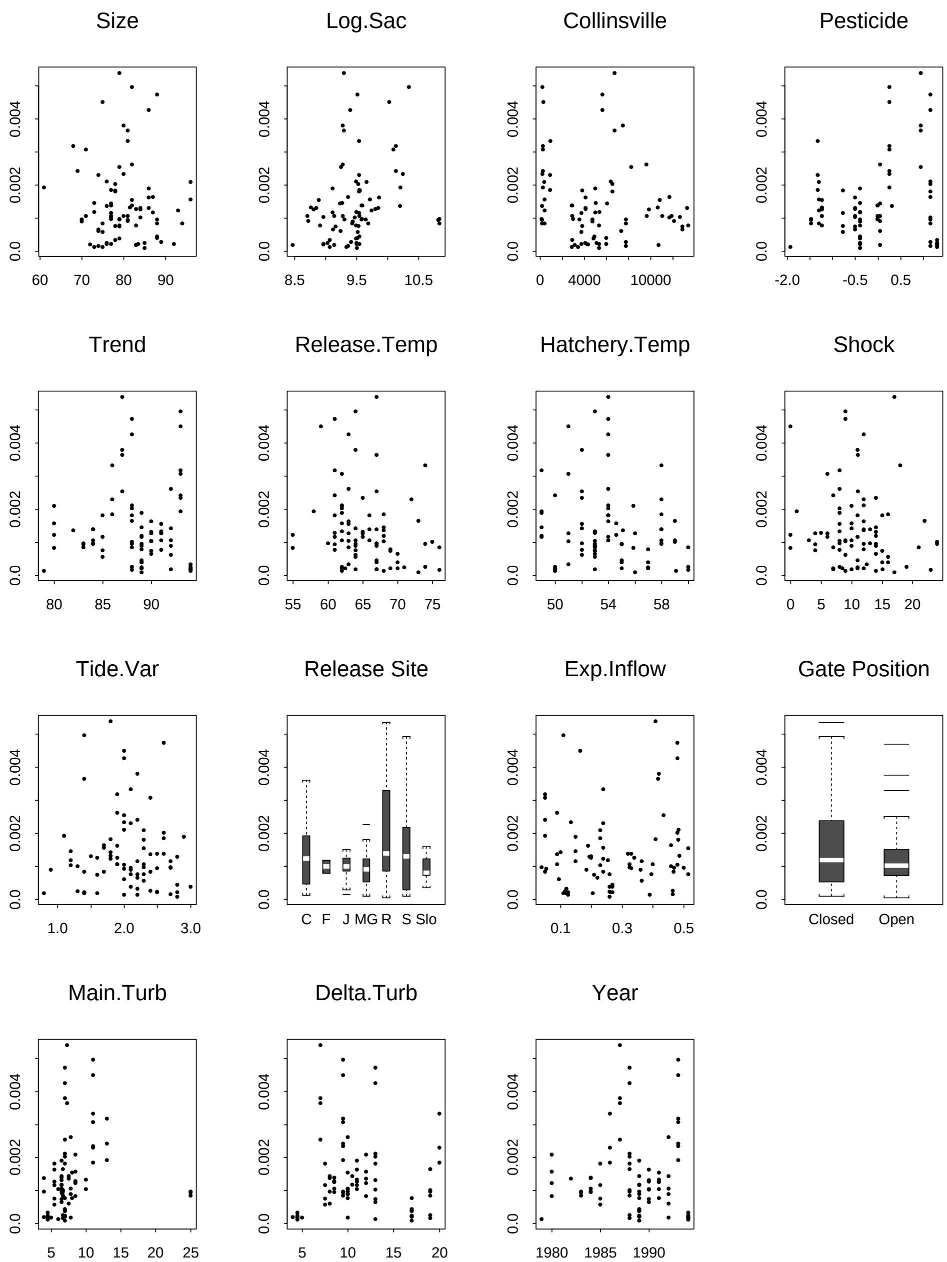


Figure 17: SI vs SI covariates, part 1. (Pesticide in standard units.)

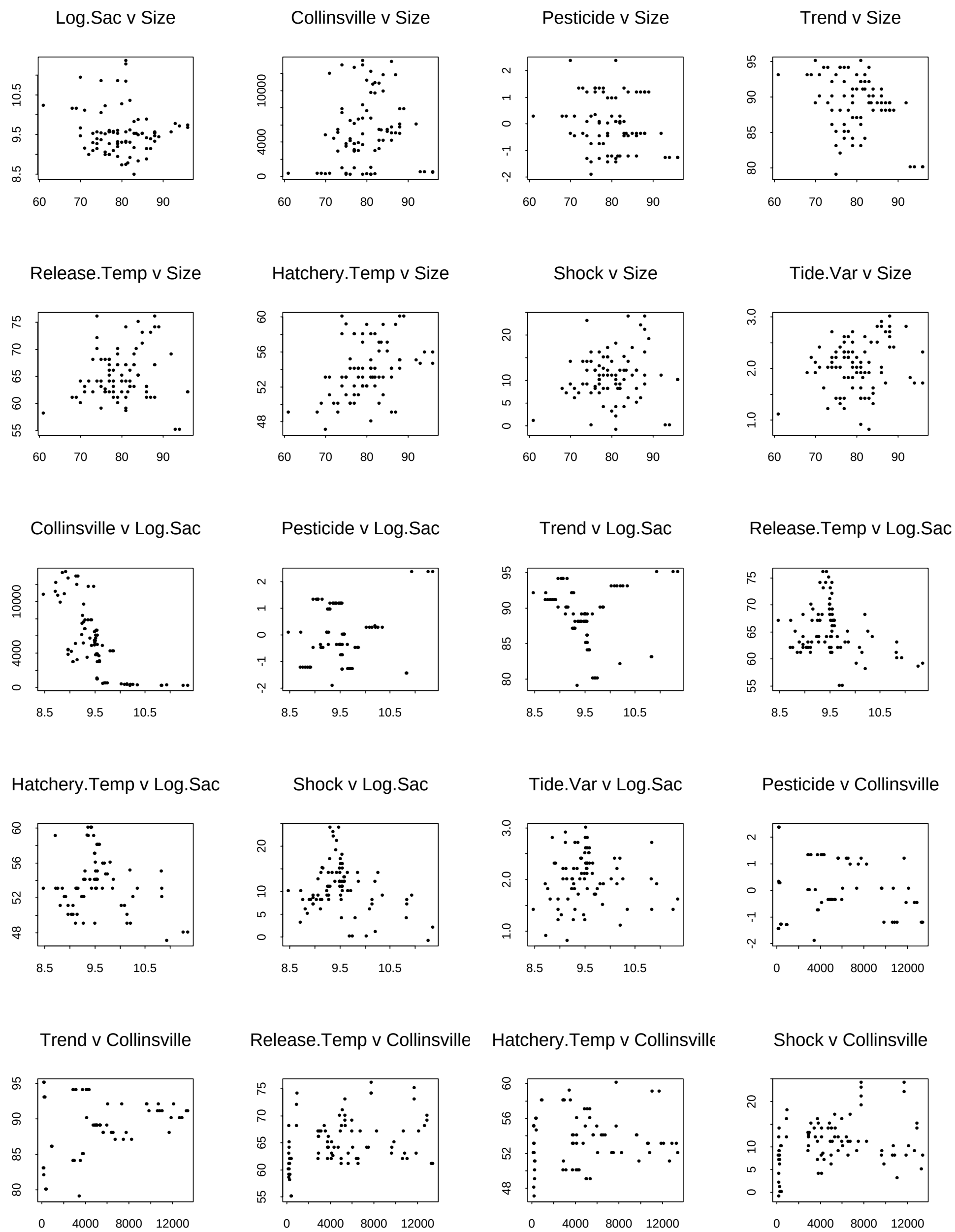


Figure 18: SI vs SI covariates, part 2. (Pesticide in standard units.)

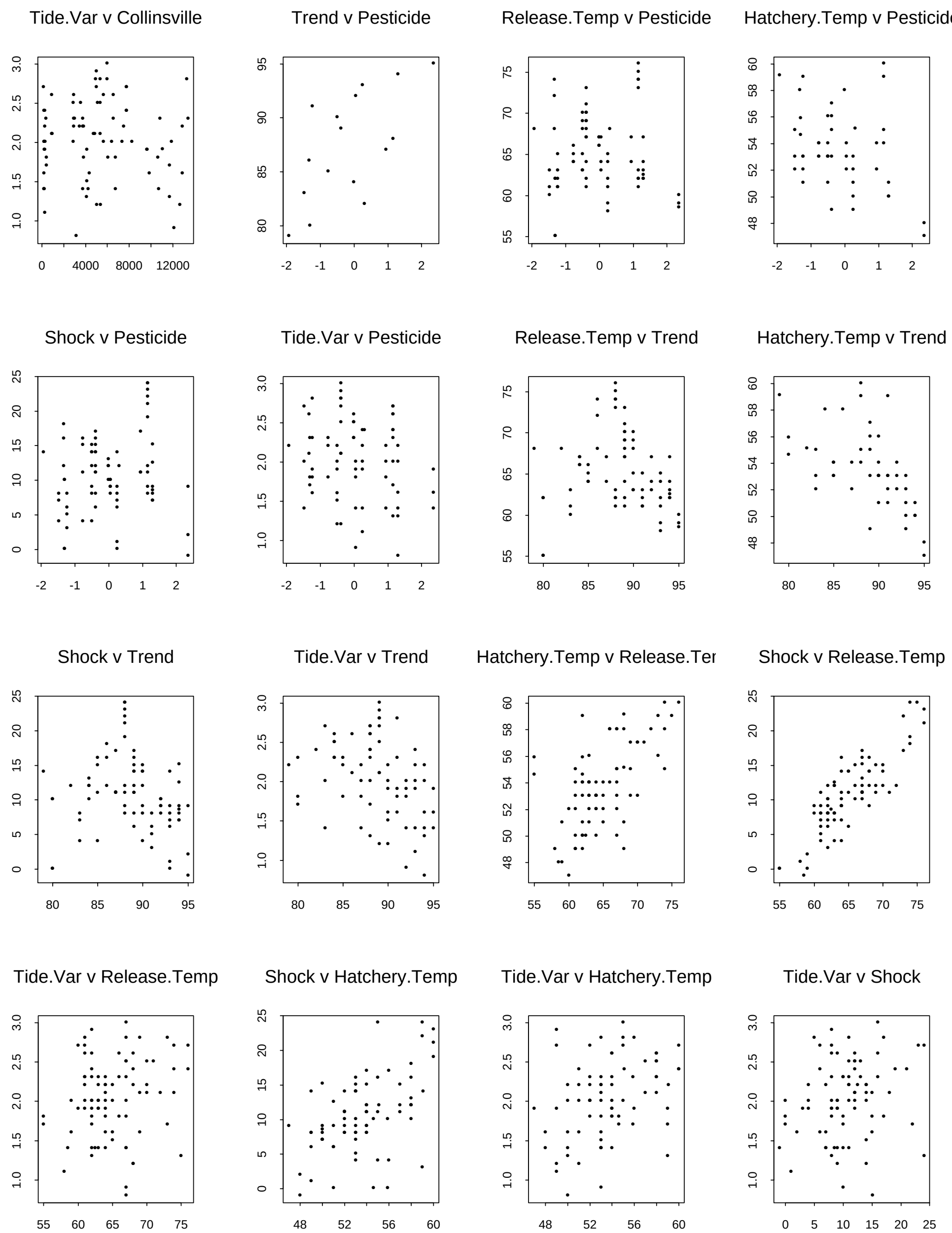


Figure 19: SD vs SD covariates.

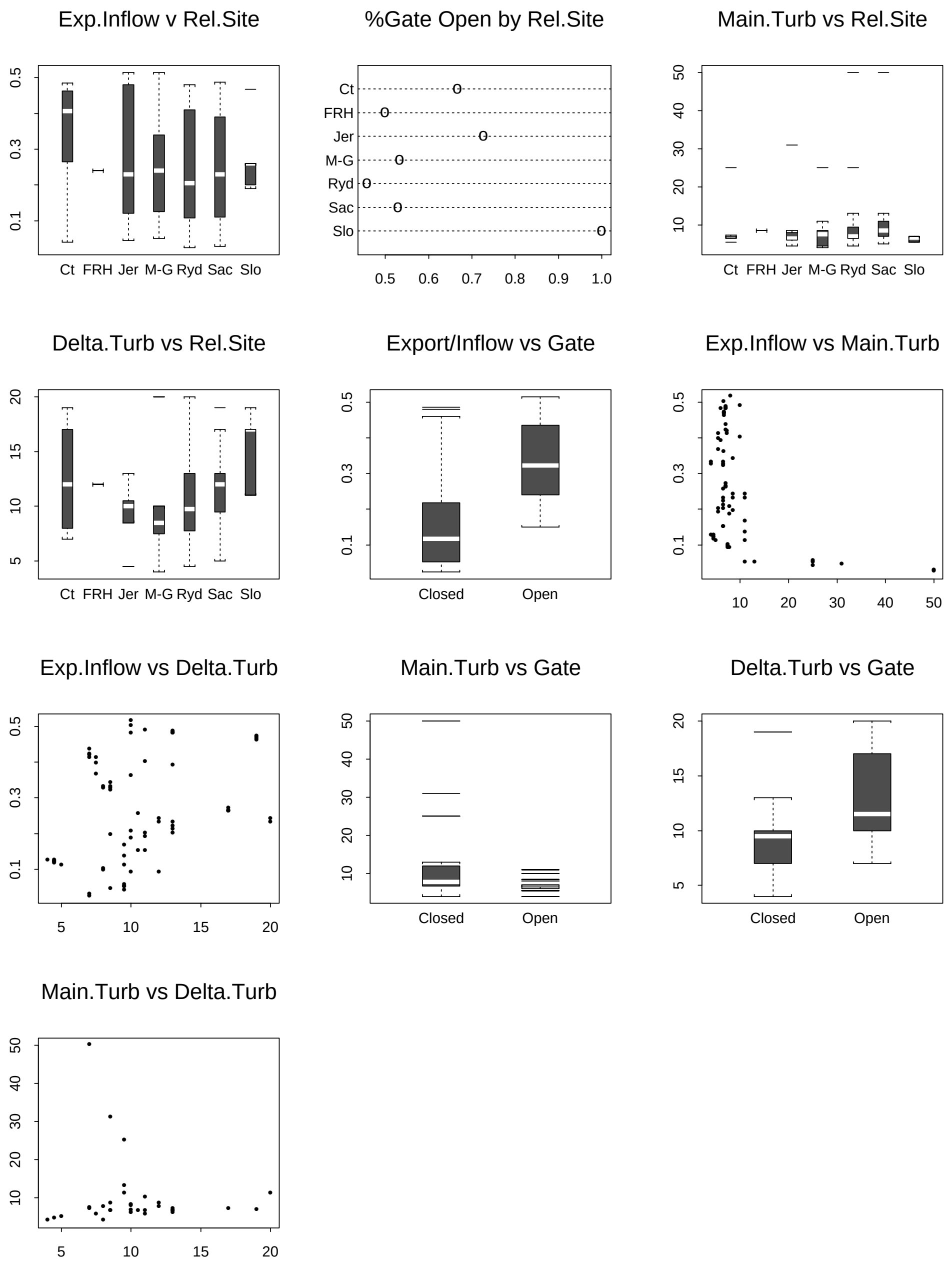


Figure 20: SI vs SD covariates, part 1.

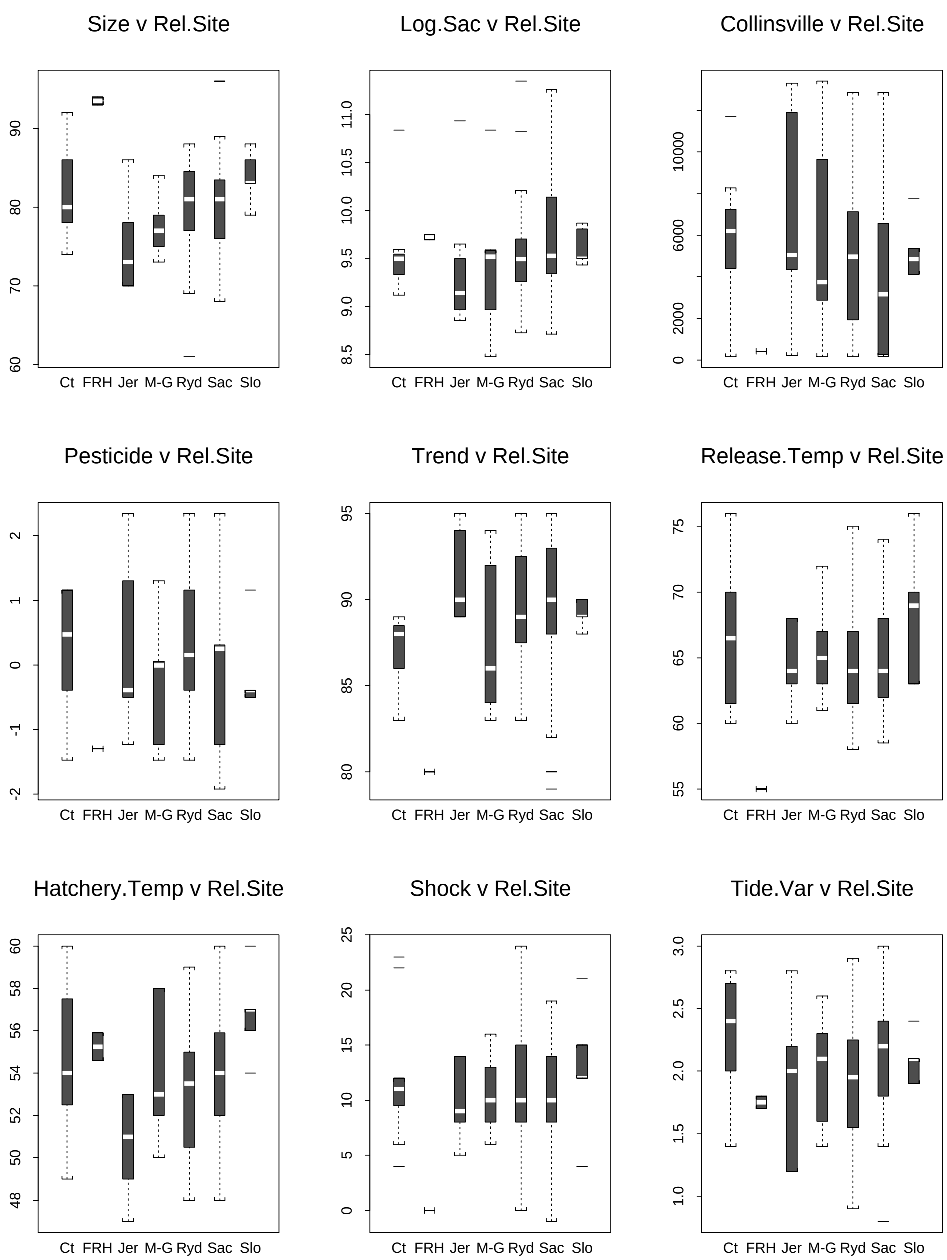


Figure 21: SI vs SD covariates, part 2.

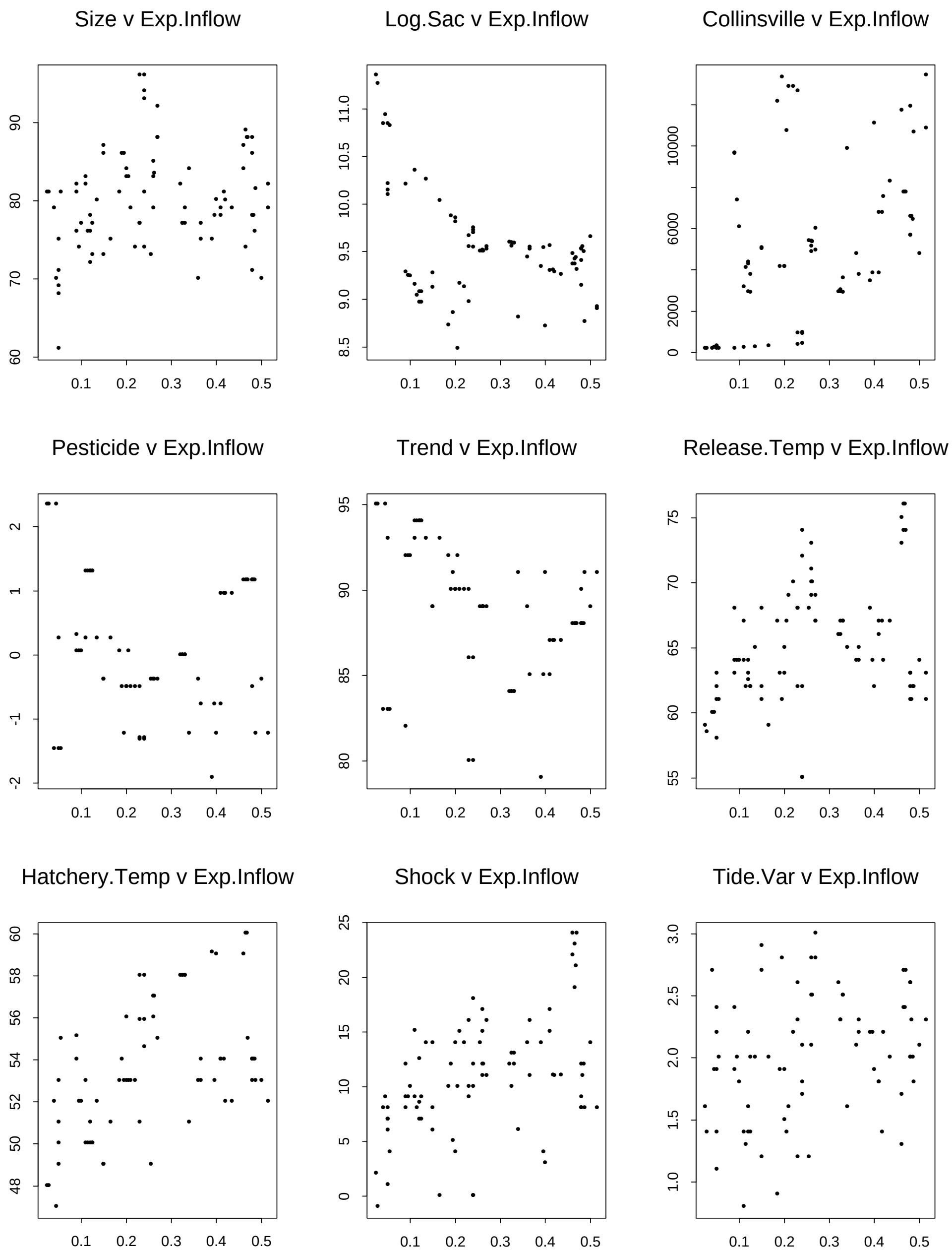


Figure 22: SI vs SD covariates, part 3.

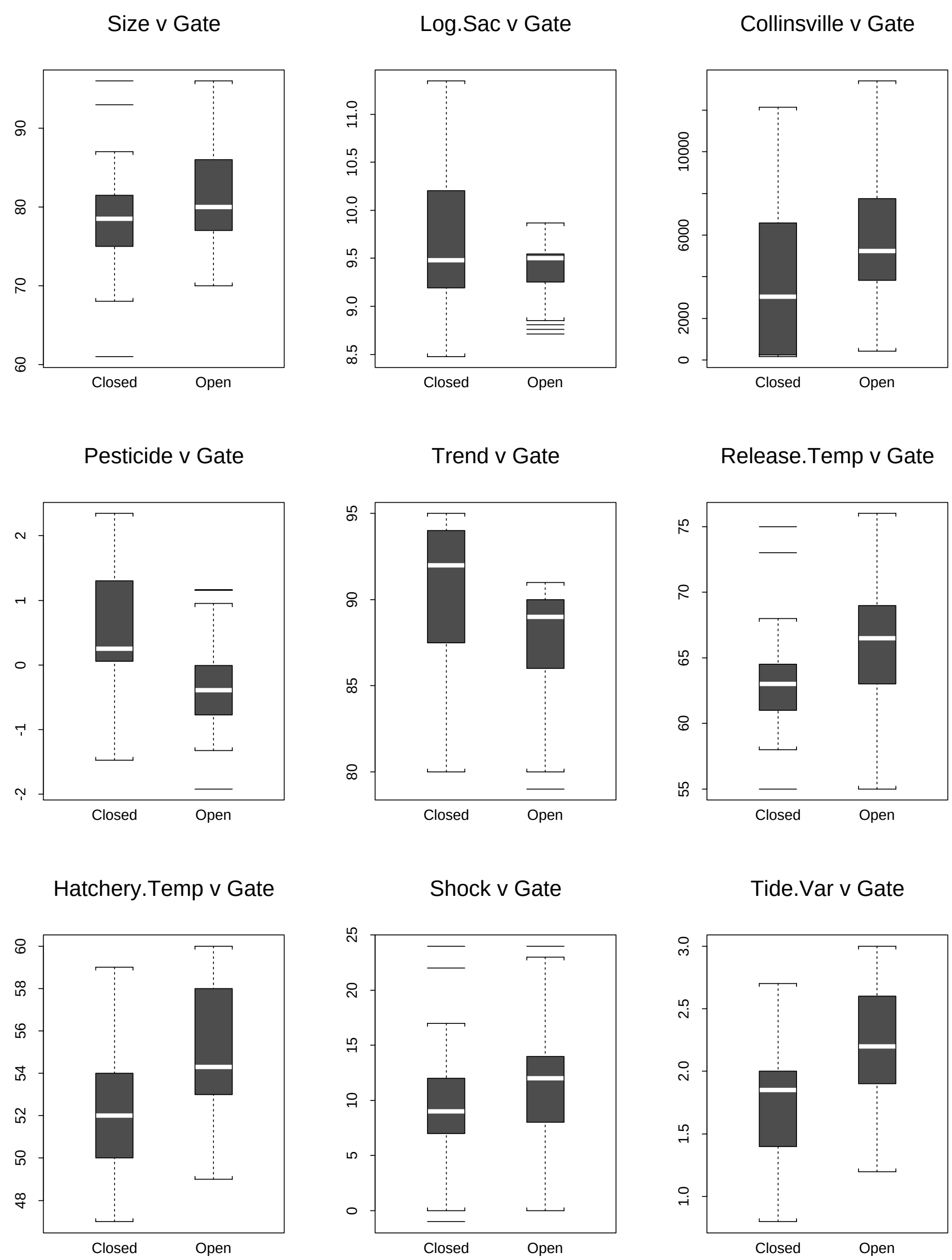


Figure 23: SI vs SD covariates, part 4.

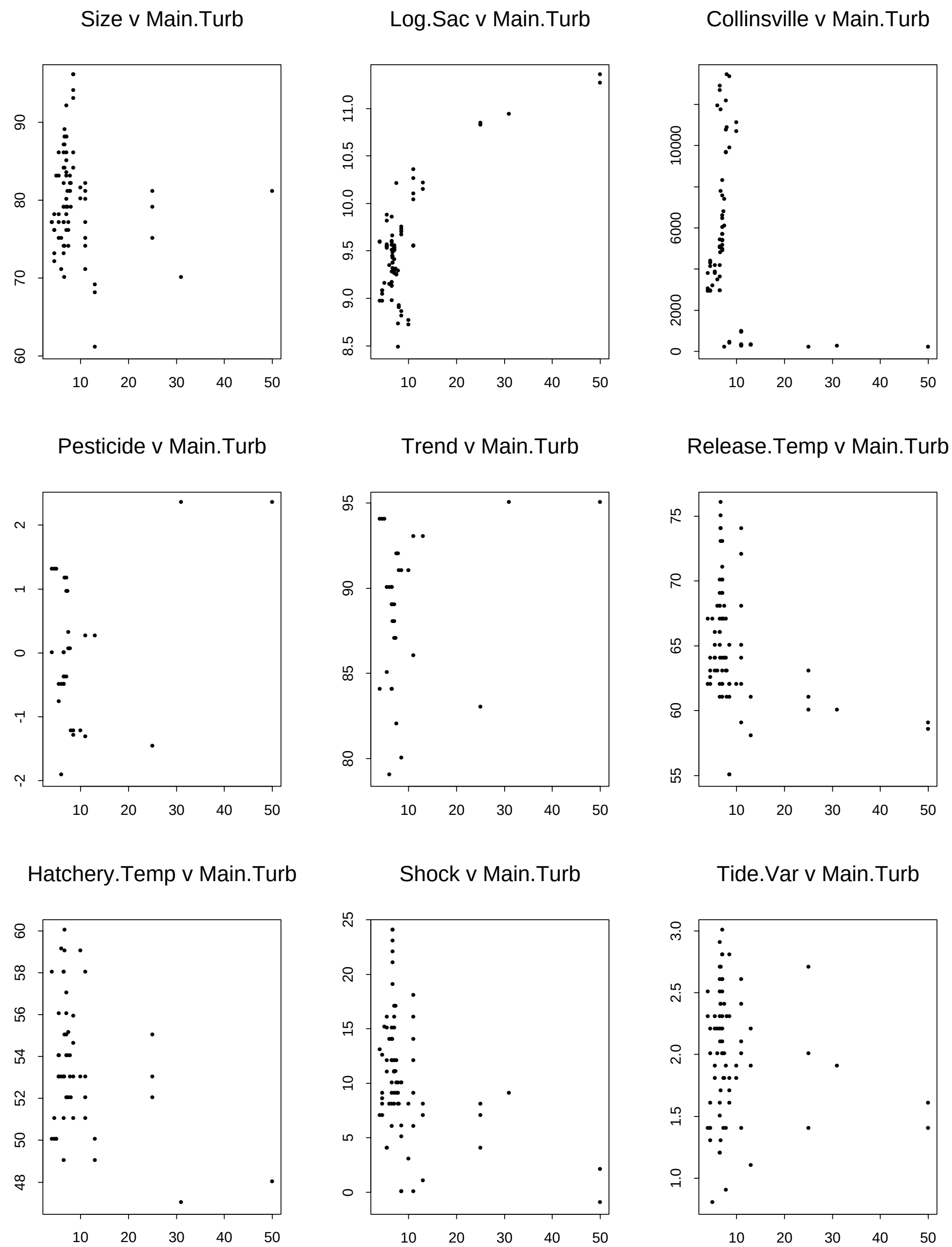
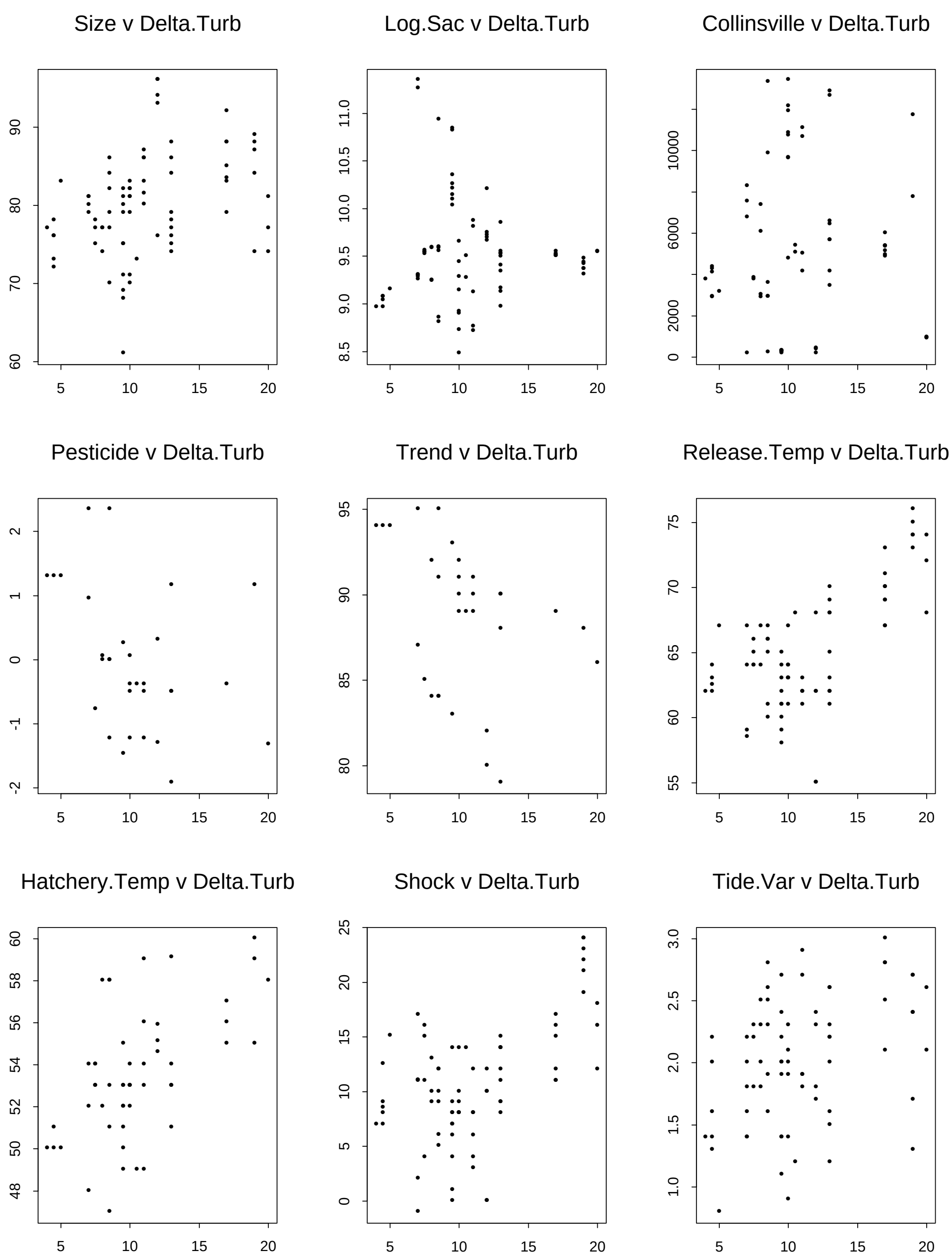


Figure 24: SI vs SD covariates, part 5.



B Model fitting details

B.1 Three aspects of parameter estimation

IWLS Similar to the least squares method used in ordinary linear regression, IWLS weights each observation inversely proportional to its variance (more variable observations get less weight) and the weights must be calculated in an iterative manner since they are functions of the coefficients to be estimated. 'Symbolically',

$$\min \sum_{i=1}^{n} \left[wt_i(\beta)(\log(Y_i) - \mathbf{x}_i^t \beta)^2 \right]$$

where $wt_i(\beta)$ is the weighting for observation i. The above is not exactly correct in that a function of $\log(Y_i)$, denoted Z_i, rather than $\log(Y_i)$ is used. In matrix terms, the resulting so-called normal equations are

$$(X^t W X)\beta \quad = \quad X^t W Z \tag{23}$$

where X is the matrix of covariates (n by p), W is a diagonal matrix of weights for each observation (n by n), β is the coefficient vector (p by 1), and Z is the vector of 'transformed' observations (n by 1).

Ridge Ridge regression is a technique for estimating regression coefficients more precisely while introducing some bias. The ridge regression penalty is a parameter λ that tends to 'stabilize' estimates of β. Namely, without its inclusion, the estimates of β may have extremely large variances, especially if some of the covariates are near linear combinations of other covariates (multicollinear). The ideal choice of λ would be one that minimizes bias and variance simultaneously. When many parameters are being estimated relative to the number of observations, as is the case here, ridge regression can be particularly valueable— non-ridge estimates can be quite variable compared to ridge estimates. If we were not trying to estimate 38 parameters with 170 observations, say, for example, only 5 or 6 parameters, the gain from using ridge would probably be negligible.

The resulting 'normal' equations for ridge estimation of GLM parameters are

$$(X^t W X + \lambda I)\beta \quad = \quad X^t W Z \tag{24}$$

where I is a p by p identity matrix. The consequent parameter estimates tend to be 'shrunk' toward zero as well.

The range of λ considered was 0 to 600. From this range, we selected λ=550 based on a combination of measures, beginning with cross-validated prediction errors. The cross-validated prediction errors were calculated using a leave-one-out approach, for each λ considered,

1. Leaving out observation i, estimate the coefficients $\beta_{-i,\lambda}$.

2. Predict the number of (Chipps Island or Ocean) recoveries for observation i, denoted $\hat{y}_{i,\lambda}$, using $\hat{\beta}_{-i,\lambda}$ and the covariate values for observation i.
3. Calculate a squared Pearson residual was calculated for each observation, namely $e^2_{i,\lambda} = (y_i - \hat{y}_{i,\lambda})^2/\hat{y}_{i,\lambda}$.
4. Various scores for the particular λ were examined, including the mean of these residuals over all observations, just the Chipps Island observations, and just the ocean observations. I.e., for $n = 170$, 86, and 84, $\frac{1}{n}\sum_{i=1}^{n} e^2_{i,\lambda}$. Trimmed means with 10% and 20% trimming were examined as well. (A z% trimmed mean is a mean calculated after removing the smaller z% and larger z% of the observations and is less sensitive to extreme values than an ordinary mean.)

Figure 25 shows the scores for the mean, 10% trimmed mean, and 20% trimmed mean over the three groupings of the observations. There are a few extremely large prediction errors for the ocean observations that greatly influence the score based on the mean for n=170 or for the ocean only observations. Emphasis was placed on prediction errors for the Chipps Island observations.

The sum of the estimated variances for the coefficients for each value of lambda are listed below.

λ	0	50	100	150	200	250	300	350	400	450	500	550	600
$\sum_{i=1}^{38} \hat{V}[\hat{\beta}_\lambda]$	0.473	0.313	0.249	0.211	0.185	0.167	0.153	0.141	0.132	0.124	0.117	0.111	0.106

Figure 26 contains several ridge *traces*, the estimated coefficients versus the size of the ridge parameter λ.

Dispersion The weight wt_i is based on the variance of observation i and the variance is $\phi E[Y]$ where ϕ is the dispersion parameter. In the model fitting process the dispersion was estimated separately for Chipps Island and ocean observations.

$$\begin{aligned}\hat{\phi}_C &= \sum_{i=1}^{86} \frac{(y_{C,i} - \hat{y}_{C,i})^2}{\hat{y}_{C,i}}/(86-p) \\ \hat{\phi}_O &= \sum_{i=1}^{84} \frac{(y_{O,i} - \hat{y}_{O,i})^2}{\hat{y}_{O,i}}/(84-p)\end{aligned}$$

where p is the number of coefficients (p=38). Estimates of ϕ_C and ϕ_O were 15.26 and 84.55. Therefore the ocean recoveries were much more variable with a dispersion parameter roughly 5.5 times larger than that for Chipps Island recoveries. Thus the weights w_i for the ocean recoveries were slightly less than 1/5th that for Chipps Island.

B.2 Standard errors

This subsection assumes standardized covariates— the means and standard deviations for each of the covariates are given in Appendix C.

Figure 25: Cross-validated prediction error as function of λ. Columns are mean, 10% trimmed mean, and 20% trimmed mean of prediction errors, while rows are those calculations based on all 170 observations, the 86 Chipps Island observations, and the 84 ocean observations.

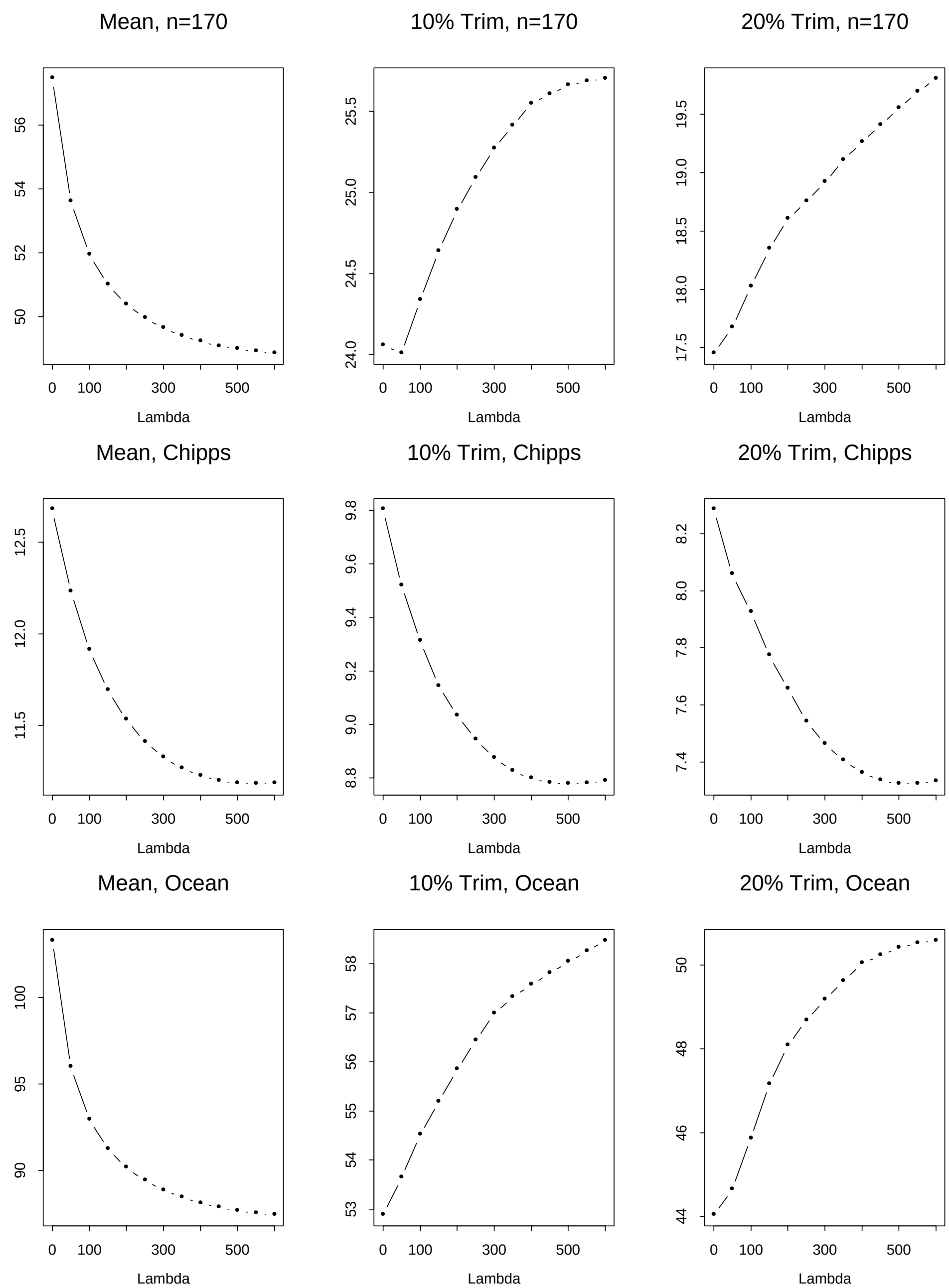
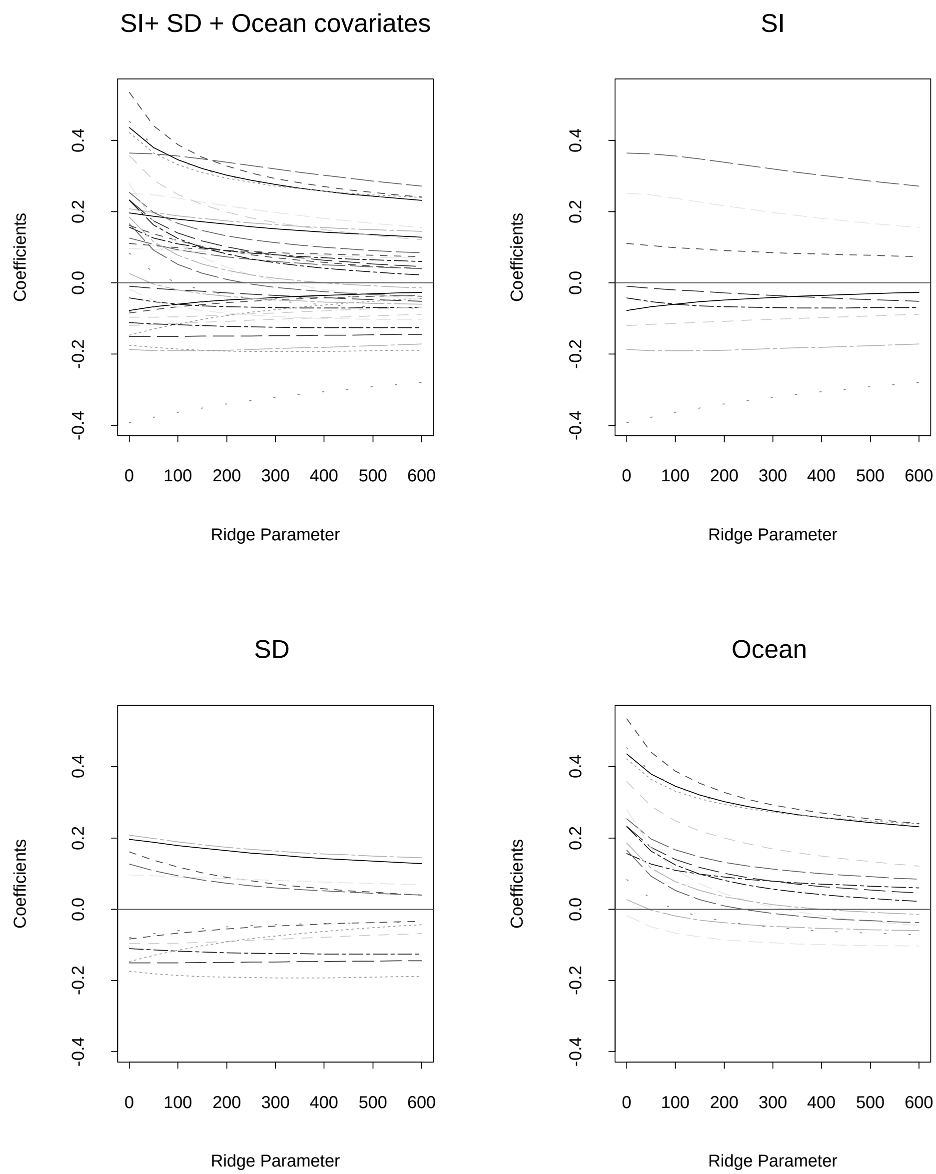


Figure 26: Ridge Traces.



To estimate the standard errors of the estimated coefficients as well as the fitted values of Y, an analytical approximation was used[12]. Formally, since

$$\hat{\beta} \;=\; (X^t W X + \lambda I)^{-1} X^t W Z$$

the variance of $\hat{\beta}$ can be shown to be approximately:

$$\mathrm{Var}[\hat{\beta}] \;\approx\; (X^t W X + \lambda I)^{-1} X^t W \mathrm{Var}[Z] W^t X (X^t W X + \lambda I)^{-1}$$

where $\mathrm{Var}[Z] = \Phi W^{-1}$, with Φ being a diagonal matrix of dispersion values.

Estimates of the expected value and the predicted value for number of recoveries given standardized covariates $x_2^*, \ldots, x_p^*$, are the same:

$$\hat{\mathrm{E}}[Y] = \hat{Y} \;=\; Rf e^{\hat{\beta}_1 + \hat{\beta}_2 x_2^* + \ldots + \hat{\beta}_p x_p^*}$$

but the variances differ because of the additional variation in predicting a particular return number as opposed to just estimating the average return number.

Let $\mathbf{x}^*$ be a p by 1 vector of standardized covariate values for a single observation. The variance of a single expected value is:

$$\begin{aligned}
\mathrm{Var}[\hat{\mathrm{E}}[Y]] &= \mathrm{Var}[Rf \exp(\mathbf{x}^* \hat{\beta})] \\
&= (Rf)^2 \mathrm{Var}[\exp(\mathbf{x}^* \hat{\beta})] \\
&\approx (Rf \exp(\mathbf{x}^* \hat{\beta}))^2 \mathbf{x}^{*t} \mathrm{Var}[\hat{\beta}] \mathbf{x}^* \\
&= \hat{\mathrm{E}}[Y]^2 \mathbf{x}^{*t} \mathrm{Var}[\hat{\beta}] \mathbf{x}^* \qquad (25)
\end{aligned}$$

The variance of a predicted value follows from the 'double variance' formula and the assumed overdispersed Poisson distribution:

$$\begin{aligned}
\mathrm{Var}[\hat{Y}] &= \mathrm{Var}_{\hat{\beta}} \mathrm{E}_{Y|\hat{\beta}}[\hat{Y}] + \mathrm{E}_{\hat{\beta}} \mathrm{Var}_{Y|\hat{\beta}}[\hat{Y}] \\
&= \mathrm{Var}_{\hat{\beta}}[\hat{\mathrm{E}}[Y]] + \mathrm{E}_{\hat{\beta}}[\hat{\phi} \hat{Y}] \qquad (26)
\end{aligned}$$

where the first term of (26) is the variance in (25).

In the case of estimating ratios of survival rates, the point estimate can be found most simply by substituting a vector of differences in covariate values and dividing the differences by the standard deviation vector (in Appendix C). The variance of the estimated expected ratio can be estimated using equation (25), substituting the differenced covariance vector divided by the standard deviation vector for $\mathbf{x}^*$ and the estimated ratio for $\hat{\mathrm{E}}[Y]$. This variance estimate is based on a first order Taylor series approximation to the ratio estimate written as a function of the estimated coefficients. A first order approximation to the variance of a prediction is to add the estimated ratio to the variance for the expected ratio (assuming equal dispersion parameters).

[12] A bootstrap procedure was originally proposed, but given the 'study design', namely fixed year effects, a simple vector resampling could not be done. The alternative bootstrapping approach, namely resampling residuals, turned out to be a problem because any reasonably defined residuals are unavoidably functions of the particular fitted values, i.e., non-pivotal.

B.3 Coefficients and error estimates on original scale.

The relationship between the coefficients on the standardized and original scales is the following:

$$\begin{aligned}\hat{\beta}_i^* &= \frac{\hat{\beta}_i}{s_i},\ i = 2, \ldots, p \\ \hat{\beta}_1^* &= \hat{\beta}_i - \sum_{i=2}^{p} \frac{\hat{\beta}_i \overline{x}_i}{s_i}\end{aligned}$$

where $\overline{x}_i$ and s_i are the mean and standard deviation of covariate x_i, β_1 is the intercept, and p is the number of covariates (in this case 39).

The corresponding variances are messy:

$$\begin{aligned}\mathrm{Var}[\hat{\beta}_i^*] &= \frac{\mathrm{Var}[\hat{\beta}_i]}{s_i^2}\ i = 2, \ldots, p \\ \mathrm{Var}[\hat{\beta}_1^*] &= \mathrm{Var}[\hat{\beta}_1] + \sum_{i=2}^{p} \frac{\overline{x}_i^2}{s_i^2} \mathrm{Var}[\hat{\beta}_i] - 2\sum_{i=2}^{p} \mathrm{Cov}[\hat{\beta}_1, \hat{\beta}_i] + 2\sum_{i=2}^{p-1} \sum_{j=i+1}^{p} \frac{\overline{x}_i \overline{x}_j}{s_i s_j} \mathrm{Cov}[\hat{\beta}_i, \hat{\beta}_j]\end{aligned}$$

As for the transformed variate case discussed in Appendix B, estimates of the expected value and the predicted value are the same, but the variances will differ. The variance for the expected value and predicted value are as in (25) and (26) with β^* substituted for β and x substituted for x^*.

Given the complexity of the variance calculations, to calculate variances of expected or predicted values, it is simpler to work with the standardized covariates and use the formulas given in (25) and (26).

C Means and standard deviations for covariates

The following means and standard deviations are based on the design matrix, a 170 by 38 matrix (excluding the intercept). Columns for the SD variables and the Year variables contain many 0's, thus the means may seem smaller than expected.

Covariate	Mean	Standard Deviation
Chipps	5.059e-01	5.014e-01
Size	8.001e+01	6.469e+00
Log.Sacramento.2	9.498e+00	5.095e-01
Collinsville	5.203e+03	3.844e+03
Pesticide	3.796e+06	6.485e+05
Trend	8.859e+01	3.873e+00
Release.Temp	6.527e+01	4.502e+00
Hatchery.Temp	5.384e+01	3.101e+00
Shock	1.087e+01	5.148e+00
Tide.Var	2.027e+00	4.945e-01
frh.dum	2.353e-02	1.520e-01
sac.dum	2.000e-01	4.012e-01
slo.dum	5.882e-02	2.360e-01
crt.dum	1.412e-01	3.492e-01
ryd.dum	2.765e-01	4.486e-01
mkg.dum	1.765e-01	3.823e-01
upper.exp.inflow	1.028e-01	1.637e-01
delta.exp.inflow	7.855e-02	1.432e-01
upper.gate	2.176e-01	4.139e-01
delta.gate	1.882e-01	3.921e-01
mainstem.turbid	5.835e+00	6.985e+00
delta.turbid	2.821e+00	4.789e+00
1979	5.882e-03	7.670e-02
1980	2.353e-02	1.520e-01
1981	5.882e-03	7.670e-02
1982	5.882e-03	7.670e-02
1983	1.765e-02	1.321e-01
1984	2.941e-02	1.695e-01
1985	2.353e-02	1.520e-01
1986	1.765e-02	1.321e-01
1987	2.353e-02	1.520e-01
1988	6.471e-02	2.467e-01
1989	8.235e-02	2.757e-01
1990	4.118e-02	1.993e-01
1991	3.529e-02	1.851e-01
1992	3.529e-02	1.851e-01
1993	4.118e-02	1.993e-01

D Recovery data: Observed and Fitted

Tables D and D show the observed and fitted values (as well as standard errors for expected and predicted values) for the Chipps Island and ocean recoveries. The standard errors for the fitted, or expected, values and the predicted values are based on equations (25) and (26). The calculation of the standard error for predicted values uses an approximation for the second term of (26) which is simply substituting the estimated values of the dispersion parameters and the fitted values.

Table 8: Chipps Island: # Released, # Observed, Fitted, Residuals, se(Fitted), se(Predicted). (Tag codes for collapsed replicates are for first code in group.)

Site (Tag Code)	Rel. #	Obs	Fit	Resid	se(fit)	se(predict)
Sac (6-62-5)	160151	50	45.4	4.6	9.3	27.9
Sac (6-62-8)	98586	33	95.9	-62.9	18.4	42.5
FRH (6-62-7)	88335	15	27.8	-12.8	8.9	22.4
Sac (6-62-11)	84643	34	64.2	-30.2	13.5	34.1
FRH (6-62-10)	88516	15	40.9	-25.9	12.8	28.1
Sac (6-62-21)	60822	30	24.0	6.0	4.6	19.7
Ct (6-62-24)	96706	92	96.1	-4.1	20.8	43.6
Moke.Georg (6-62-25)	83435	72	45.2	26.8	11.0	28.5
Ryde (6-62-23)	92693	95	102.3	-7.3	19.0	43.8
Ct (6-62-27)	62604	37	18.0	19.0	2.8	16.8
Moke.Georg (6-42-8)	14916	9	4.8	4.2	0.9	8.6
Moke.Georg (6-62-28)	41371	25	11.8	13.2	2.2	13.6
Ryde (6-62-29)	44818	30	28.0	2.0	4.2	21.1
Moke.Georg (6-62-32)	59808	24	15.8	8.2	2.9	15.8
Moke.Georg (6-62-34)	100386	25	41.3	-16.3	8.4	26.5
Moke.Georg (6-62-36)	101236	30	56.4	-26.4	9.4	30.8
Ct (6-62-38)	100626	37	59.5	-22.5	9.0	31.4
Ryde (6-62-35)	107161	88	95.5	-7.5	11.6	39.9
Moke.Georg (6-62-46)	102965	24	28.7	-4.7	6.7	22.0
Moke.Georg (6-62-47)	101949	33	22.1	10.9	5.3	19.1
Ryde (6-62-48)	101320	74	40.9	33.1	7.4	26.0
Ct (6-62-53)	100302	71	54.7	16.3	7.9	29.9
Ryde (6-62-55)	51103	46	33.8	12.2	4.0	23.1
Ct (6-62-56)	100919	43	42.0	1.0	6.2	26.1
Ryde (6-62-58)	51008	47	46.7	0.3	5.7	27.3
Ct (B6-14-2)	107249	159	147.4	11.6	21.8	52.2
Ryde (6-31-1)	52741	105	91.6	13.4	10.8	38.9
Sac (B6-14-6)	102736	142	97.7	44.3	16.0	41.8
Ct (B6-14-4)	102480	147	105.0	42.0	15.4	42.9
Ryde (6-31-2)	53238	145	111.1	33.9	15.2	43.9
Ct (6-62-59)	106901	39	91.1	-52.1	15.9	40.5
Ryde (6-62-63)	53961	46	54.4	-8.4	9.8	30.4
Sac (6-62-61)	97892	14	38.7	-24.7	7.1	25.3
Ct (6-62-50)	99827	5	33.5	-28.5	7.0	23.6
Sloughs (6-31-5)	97317	79	68.5	10.5	14.2	35.3
Ryde (6-31-3)	53942	39	44.7	-5.7	7.5	27.2
Jers (6-1-11-1-11)	27758	24	19.1	4.9	3.3	17.4
Jers (6-1-11-1-12)	29058	29	19.8	9.2	4.0	17.8
Ct (6-31-11)	51211	46	38.0	8.0	7.4	25.2
Ryde (6-31-12)	51046	65	60.9	4.1	9.7	32.0
Jers (6-1-11-1-10)	28708	25	20.9	4.1	4.1	18.3
Jers (6-1-11-1-9)	27525	33	19.1	13.9	3.9	17.5
Sac (6-31-10)	52612	9	23.6	-14.6	3.7	19.3
Ct (6-31-8)	50659	19	22.8	-3.8	4.1	19.1
Ryde (6-31-7)	50601	26	42.3	-16.3	5.5	26.0
Sloughs (6-1-14-1-1)	51237	50	30.1	19.9	5.6	22.2
Sloughs (6-31-16)	49762	57	33.2	23.8	6.0	23.3
Sac (6-31-15)	94604	20	31.9	-11.9	4.5	22.5
Ct (6-1-14-1-3)	52907	17	17.2	-0.2	3.0	16.5
Ryde (6-1-14-1-2)	51134	8	24.1	-16.1	3.7	19.5
Jers (6-1-14-1-9)	52962	32	55.9	-23.9	10.4	31.0
Jers (6-31-19)	50143	56	49.7	6.3	9.6	29.2
Sac (6-31-18)	48390	44	20.8	23.2	4.0	18.3
Ryde (6-31-20)	51878	89	44.8	44.2	6.6	27.0
Sloughs (6-31-23)	49324	39	56.5	-17.5	11.2	31.4
Sloughs (6-31-21)	52010	58	62.2	-4.2	12.4	33.2
Ryde (6-31-22)	50837	67	52.6	14.4	6.9	29.2
Moke.Georg (6-1-14-2-5)	47289	79	56.2	22.8	11.3	31.4
Jers (6-1-14-2-6)	52139	94	64.3	29.7	12.4	33.7
Sac (6-1-14-2-7)	102664	85	83.5	1.5	15.2	38.8
Sac (6-1-14-2-9)	104516	55	75.1	-20.1	15.7	37.3
Moke.Georg (6-31-27)	45706	31	39.8	-8.8	7.6	25.8
Jers (6-31-28)	49184	89	98.6	-9.6	18.8	43.1
Moke.Georg (6-1-14-2-10)	51846	23	23.7	-0.7	4.1	19.4
Ryde (6-1-14-2-11)	53630	78	47.7	30.3	4.9	27.4
Moke.Georg (6-1-14-3-2)	52374	41	29.5	11.5	5.5	21.9
Ryde (6-1-14-3-1)	42534	97	53.2	43.8	6.2	29.2
Moke.Georg (6-31-30)	51914	11	20.4	-9.4	4.3	18.1
Ryde (6-31-29)	53099	93	54.0	39.0	8.7	30.0
Ryde (6-1-14-3-9)	53265	23	57.0	-34.0	10.8	31.4
Sac (6-1-14-3-11)	54454	36	36.3	-0.3	6.8	24.5
Ryde (6-1-14-11-3)	28056	19	21.7	-2.7	3.3	18.5
Sac (6-1-14-3-13)	51574	23	30.8	-7.8	5.7	22.4
Ryde (6-31-37)	49699	43	48.8	-5.8	7.3	28.2
Sac (6-31-39)	49786	18	25.8	-7.8	4.5	20.3
Sac (6-31-40)	50116	40	31.8	8.2	5.7	22.7
Moke.Georg (6-1-14-4-2)	51485	3	18.8	-15.8	3.8	17.3
Ryde (6-1-14-4-1)	51819	11	34.5	-23.5	5.1	23.5
Jers (6-1-14-4-3)	50689	10	20.7	-10.7	3.7	18.2
Moke.Georg (6-1-14-4-7)	50235	6	14.9	-8.9	2.9	15.4
Ryde (6-1-14-4-6)	56139	11	40.4	-29.4	6.0	25.5
Jers (6-1-14-4-8)	53810	16	20.2	-4.2	3.5	17.9
Sac (6-31-42)	53232	4	21.7	-17.7	4.6	18.8
Jers (6-1-14-4-13)	50779	25	30.3	-5.3	8.3	23.1
Sac (6-1-14-5-3)	50292	32	37.0	-5.0	12.9	27.1
Ryde (6-1-14-5-2)	51597	46	46.2	-0.2	15.5	30.7

Table 9: Ocean: # Released, # Observed, Fitted, Residuals, se(Fitted), se(Predicted). (Tag codes for collapsed replicates are for first code in group.)

Site (Tag Code)	Rel. #	Obs	Fit	Resid	se(fit)	se(predict)
Sac (6-62-5)	160151	98	122.0	-24.0	88.8	134.9
Sac (6-62-8)	98586	951	986.7	-35.7	207.2	355.5
FRH (6-62-7)	88335	292	247.0	45.0	73.4	162.1
Sac (6-62-11)	84643	678	728.0	-50.0	159.5	294.9
FRH (6-62-10)	88516	497	415.8	81.2	120.3	222.8
Sac (6-62-14)	140250	47	60.7	-13.7	55.7	90.8
Sac (6-62-21)	60822	392	377.4	14.6	170.7	247.1
Ct (6-62-24)	96706	428	443.0	-15.0	127.8	231.9
Moke.Georg (6-62-25)	83435	270	186.1	83.9	61.4	139.7
Ryde (6-62-23)	92693	369	465.9	-96.9	130.2	237.4
Ct (6-62-27)	62604	400	274.7	125.3	76.2	170.4
Moke.Georg (6-42-8)	14916	61	61.9	-0.9	16.7	74.3
Moke.Georg (6-62-28)	41371	289	198.4	90.6	53.6	140.2
Ryde (6-62-29)	44818	194	318.3	-124.3	85.8	185.1
Moke.Georg (6-62-32)	59808	266	291.2	-25.2	78.4	175.4
Moke.Georg (6-62-34)	100386	317	400.4	-83.4	97.0	208.0
Moke.Georg (6-62-36)	101236	565	471.9	93.1	106.0	226.1
Ct (6-62-38)	100626	387	527.6	-140.6	120.1	243.0
Ryde (6-62-35)	107161	924	770.3	153.7	160.0	301.2
Moke.Georg (6-62-46)	102965	984	1311.7	-327.7	224.8	401.8
Moke.Georg (6-62-47)	101949	1306	1108.8	197.2	186.7	358.6
Ryde (6-62-48)	101320	1978	1752.2	225.8	311.7	495.3
Ct (6-62-53)	100302	1983	1785.9	197.1	263.2	469.3
Ryde (6-62-55)	51103	1610	1186.0	424.0	168.8	358.9
Ct (6-62-56)	100919	1435	1440.4	-5.4	212.4	408.5
Ryde (6-62-58)	51008	1037	1540.2	-503.2	215.3	420.2
Ct (B6-14-2)	107249	1190	1227.5	-37.5	181.0	369.5
Ryde (6-31-1)	52741	1075	724.8	350.2	97.3	266.0
Sac (B6-14-6)	102736	1118	878.6	239.4	146.8	309.6
Ct (B6-14-4)	102480	936	886.2	49.8	128.2	302.3
Ryde (6-31-2)	53238	1322	962.3	359.7	140.2	317.8
Ct (6-62-59)	106901	1037	890.7	146.3	142.5	309.2
Ryde (6-62-63)	53961	252	426.2	-174.2	76.7	204.8
Sac (6-62-61)	97892	141	404.9	-263.9	70.7	198.1
Ct (6-62-50)	99827	70	252.7	-182.7	49.7	154.4
Sloughs (6-31-5)	97317	390	555.5	-165.5	115.7	245.7
Ryde (6-31-3)	53942	285	421.1	-136.1	70.5	201.4
Jers (6-1-11-1-11)	27758	83	68.9	14.1	17.7	78.4
Jers (6-1-11-1-12)	29058	97	74.8	22.2	20.9	82.2
Ct (6-31-11)	51211	246	168.1	77.9	42.7	126.6
Ryde (6-31-12)	51046	417	281.4	135.6	66.0	167.8
Jers (6-1-11-1-10)	28708	139	92.4	46.6	25.0	91.9
Jers (6-1-11-1-9)	27525	144	74.2	69.8	20.4	81.8
Sac (6-31-10)	52612	55	71.4	-16.4	17.1	79.5
Ct (6-31-8)	50659	41	108.3	-67.3	26.7	99.3
Ryde (6-31-7)	50601	82	172.3	-90.3	38.5	126.7
Sloughs (6-1-14-1-1)	51237	71	123.5	-52.5	32.1	107.1
Sloughs (6-31-16)	49762	151	148.0	3.0	38.0	118.1
Sac (6-31-15)	94604	68	133.2	-65.2	31.1	110.6
Ct (6-1-14-1-3)	52907	47	76.6	-29.6	18.8	82.7
Ryde (6-1-14-1-2)	51134	10	91.7	-81.7	21.3	90.6
Jers (6-1-14-1-9)	52962	224	253.9	-29.9	68.2	161.6
Jers (6-31-19)	50143	208	229.0	-21.0	60.9	151.9
Sac (6-31-18)	48390	115	99.0	16.0	27.7	95.6
Ryde (6-31-20)	51878	173	281.7	-108.7	70.1	169.5
Sloughs (6-31-23)	49324	221	210.3	10.7	57.6	145.2
Sloughs (6-31-21)	52010	319	244.0	75.0	66.3	158.2
Ryde (6-31-22)	50837	218	181.3	36.7	45.7	132.0
Moke.Georg (6-1-14-2-5)	47289	143	260.5	-117.5	61.7	160.7
Jers (6-1-14-2-6)	52139	356	331.7	24.3	78.5	184.9
Sac (6-1-14-2-7)	102664	600	427.1	172.9	98.2	213.9
Sac (6-1-14-2-9)	104516	463	371.5	91.5	91.0	199.2
Moke.Georg (6-31-27)	45706	231	188.1	42.9	45.7	134.2
Jers (6-31-28)	49184	274	494.5	-220.5	113.4	233.8
Moke.Georg (6-1-14-2-10)	51846	144	160.3	-16.3	40.9	123.4
Ryde (6-1-14-2-11)	53630	350	310.1	39.9	73.2	177.7
Moke.Georg (6-1-14-3-2)	52374	240	182.0	58.0	47.2	132.7
Ryde (6-1-14-3-1)	42534	490	331.2	158.8	78.0	184.6
Moke.Georg (6-31-30)	51914	31	110.9	-79.9	29.7	101.3
Ryde (6-31-29)	53099	218	363.7	-145.7	88.8	196.6
Ryde (6-1-14-3-9)	53265	400	916.5	-516.5	163.6	322.9
Sac (6-1-14-3-11)	54454	645	557.4	87.6	95.5	237.2
Ryde (6-1-14-11-3)	28056	263	366.1	-103.1	59.9	185.9
Sac (6-1-14-3-13)	51574	611	494.3	116.7	84.9	221.4
Ryde (6-31-37)	49699	872	815.0	57.0	137.2	296.2
Sac (6-31-39)	49786	451	418.8	32.2	75.7	202.8
Sac (6-31-40)	50116	975	504.3	470.7	92.6	226.3
Moke.Georg (6-1-14-4-2)	51485	30	69.3	-39.3	15.5	78.1
Ryde (6-1-14-4-1)	51819	37	172.6	-135.6	30.4	124.6
Jers (6-1-14-4-3)	50689	32	97.6	-65.6	19.8	93.0
Moke.Georg (6-1-14-4-7)	50235	23	81.7	-58.7	17.8	85.0
Ryde (6-1-14-4-6)	56139	51	203.8	-152.8	36.1	136.1
Jers (6-1-14-4-8)	53810	66	101.2	-35.2	20.6	94.8
Sac (6-31-42)	53232	36	117.8	-81.8	27.2	103.4

E Covariance matrix for estimated coefficients

The entire 38 by 38 covariance matrix for the standardized coefficients can be retrieved at the URL
http://www.uidaho.edu/~newman/usfwsglm.html
(stored as component labelled `cov.beta` in S-PLUS list object `temp.550`).

The 21 by 21 submatrix for the SI and SD covariates is given here.

	Size	Log.Sacramento.2	Collinsville	Pesticide	Trend	Release.Temp	Hatchery.Temp
Size	2.58e-03	3.58e-04	-2.84e-04	1.25e-04	5.28e-04	-1.02e-04	-3.89e-04
Log.Sacramento.2	3.58e-04	4.09e-03	1.55e-03	-1.69e-04	5.43e-04	9.77e-05	-1.10e-04
Collinsville	-2.84e-04	1.55e-03	3.81e-03	5.18e-05	-8.75e-04	-4.51e-04	-3.57e-04
Pesticide	1.25e-04	-1.69e-04	5.18e-05	3.01e-03	-1.24e-03	-1.25e-05	3.12e-04
Trend	5.28e-04	5.43e-04	-8.75e-04	-1.24e-03	4.18e-03	-1.50e-04	9.46e-04
Release.Temp	-1.02e-04	9.77e-05	-4.51e-04	-1.25e-05	-1.50e-04	3.95e-03	-1.19e-03
Hatchery.Temp	-3.89e-04	-1.10e-04	-3.57e-04	3.12e-04	9.46e-04	-1.19e-03	3.94e-03
Shock	-1.13e-04	2.24e-07	-1.89e-04	-5.91e-04	4.63e-04	-2.07e-03	-4.91e-04
Tide.Var	-4.32e-04	-8.13e-05	1.01e-04	-3.53e-05	2.96e-04	4.50e-04	-2.04e-04
frh.dum	-3.22e-04	1.78e-06	6.86e-05	-2.15e-05	3.80e-04	3.57e-04	-2.87e-04
sac.dum	-2.88e-04	-1.05e-04	2.67e-04	1.98e-04	-1.33e-04	-1.04e-04	1.31e-05
slo.dum	-1.55e-04	-4.52e-04	9.93e-05	1.35e-05	-3.96e-04	-2.98e-04	-6.53e-04
crt.dum	3.64e-05	-3.06e-04	-1.50e-04	-1.06e-04	2.42e-04	-1.71e-04	3.47e-04
ryd.dum	1.40e-04	2.43e-04	-2.69e-04	-1.39e-04	2.06e-04	-1.97e-06	-5.82e-05
mkg.dum	8.09e-05	-1.47e-04	3.42e-04	-1.05e-04	9.83e-05	7.66e-05	-2.49e-04
upper.exp.inflow	3.19e-04	4.08e-04	-4.07e-04	-3.98e-04	3.47e-04	3.31e-04	-7.89e-04
delta.exp.inflow	3.47e-04	1.56e-04	-5.18e-04	2.90e-04	4.15e-05	8.00e-05	2.74e-04
upper.gate	2.32e-04	3.08e-04	-1.42e-04	1.91e-04	1.08e-04	-2.57e-04	2.15e-05
delta.gate	7.44e-05	4.19e-05	-5.19e-04	3.15e-04	9.32e-05	-5.09e-05	-5.14e-05
mainstem.turbid	-1.99e-04	-1.61e-03	-7.97e-06	-1.14e-05	-2.58e-04	-8.87e-06	2.71e-05
delta.turbid	3.06e-04	-2.86e-04	-9.65e-05	2.27e-05	-4.72e-05	-8.23e-05	1.54e-05

	Shock	Tide.Var	frh.dum	sac.dum	slo.dum	crt.dum	ryd.dum	mkg.dum
Size	-1.13e-04	-4.32e-04	-3.22e-04	-2.88e-04	-1.55e-04	3.64e-05	1.40e-04	8.09e-05
Log.Sacramento.2	2.24e-07	-8.13e-05	1.78e-06	-1.05e-04	-4.52e-04	-3.06e-04	2.43e-04	-1.47e-04
Collinsville	-1.89e-04	1.01e-04	6.86e-05	2.67e-04	9.93e-05	-1.50e-04	-2.69e-04	3.42e-04
Pesticide	-5.91e-04	-3.53e-05	-2.15e-05	1.98e-04	1.35e-05	-1.06e-04	-1.39e-04	-1.05e-04
Trend	4.63e-04	2.96e-04	3.80e-04	-1.33e-04	-3.96e-04	2.42e-04	2.06e-04	9.83e-05
Release.Temp	-2.07e-03	4.50e-04	3.57e-04	-1.04e-04	-2.98e-04	-1.71e-04	-1.97e-06	7.66e-05
Hatchery.Temp	-4.91e-04	-2.04e-04	-2.87e-04	1.31e-05	-6.53e-04	3.47e-04	-5.82e-05	-2.49e-04
Shock	3.59e-03	-1.40e-04	4.16e-04	-1.20e-04	3.20e-04	-7.59e-05	7.99e-05	8.85e-06
Tide.Var	-1.40e-04	2.10e-03	2.91e-04	-4.03e-05	-4.54e-05	-1.18e-04	4.07e-05	7.62e-05
frh.dum	4.16e-04	2.91e-04	2.33e-03	5.39e-04	6.90e-05	2.58e-04	1.72e-04	1.22e-04
sac.dum	-1.20e-04	-4.03e-05	5.39e-04	2.68e-03	2.60e-04	1.16e-03	6.69e-04	1.87e-04
slo.dum	3.20e-04	-4.54e-05	6.90e-05	2.60e-04	2.43e-03	1.19e-04	7.95e-04	3.00e-04
crt.dum	-7.59e-05	-1.18e-04	2.58e-04	1.16e-03	1.19e-04	2.48e-03	6.01e-04	2.31e-04
ryd.dum	7.99e-05	4.07e-05	1.72e-04	6.69e-04	7.95e-04	6.01e-04	2.86e-03	5.47e-04
mkg.dum	8.85e-06	7.62e-05	1.22e-04	1.87e-04	3.00e-04	2.31e-04	5.47e-04	3.63e-03
upper.exp.inflow	1.85e-04	3.29e-04	-1.48e-04	-8.02e-04	6.52e-04	-1.50e-03	1.27e-03	2.30e-04
delta.exp.inflow	-1.25e-04	-2.14e-04	-7.74e-05	-7.62e-06	2.06e-04	-5.10e-06	4.17e-04	-6.01e-04
upper.gate	4.82e-04	-4.99e-04	-6.21e-05	-5.02e-04	1.26e-04	3.22e-05	1.87e-04	2.40e-05
delta.gate	2.72e-04	-1.71e-04	1.64e-04	3.64e-04	3.95e-04	2.40e-04	6.82e-04	1.97e-04
mainstem.turbid	4.38e-04	1.21e-04	8.98e-05	-3.11e-04	7.43e-04	-2.02e-04	-3.30e-04	4.49e-04
delta.turbid	2.27e-04	5.43e-06	-7.86e-06	-3.63e-05	3.36e-04	2.54e-06	5.57e-04	-1.23e-03

	upper.exp.inflow	delta.exp.inflow	upper.gate	delta.gate	mainstem.turbid	delta.turbid
Size	3.19e-04	3.47e-04	2.32e-04	7.44e-05	-1.99e-04	3.06e-04
Log.Sacramento.2	4.08e-04	1.56e-04	3.08e-04	4.19e-05	-1.61e-03	-2.86e-04
Collinsville	-4.07e-04	-5.18e-04	-1.42e-04	-5.19e-04	-7.97e-06	-9.65e-05
Pesticide	-3.98e-04	2.90e-04	1.91e-04	3.15e-04	-1.14e-05	2.27e-05
Trend	3.47e-04	4.15e-05	1.08e-04	9.32e-05	-2.58e-04	-4.72e-05
Release.Temp	3.31e-04	8.00e-05	-2.57e-04	-5.09e-05	-8.87e-06	-8.23e-05
Hatchery.Temp	-7.89e-04	2.74e-04	2.15e-05	-5.14e-05	2.71e-05	1.54e-05
Shock	1.85e-04	-1.25e-04	4.82e-04	2.72e-04	4.38e-04	2.27e-04
Tide.Var	3.29e-04	-2.14e-04	-4.99e-04	-1.71e-04	1.21e-04	5.43e-06
frh.dum	-1.48e-04	-7.74e-05	-6.21e-05	1.64e-04	8.98e-05	-7.86e-06
sac.dum	-8.02e-04	-7.62e-06	-5.02e-04	3.64e-04	-3.11e-04	-3.63e-05
slo.dum	6.52e-04	2.06e-04	1.26e-04	3.95e-04	7.43e-04	3.36e-04
crt.dum	-1.50e-03	-5.10e-06	3.22e-05	2.40e-04	-2.02e-04	2.54e-06
ryd.dum	1.27e-03	4.17e-04	1.87e-04	6.82e-04	-3.30e-04	5.57e-04
mkg.dum	2.30e-04	-6.01e-04	2.40e-05	1.97e-04	4.49e-04	-1.23e-03
upper.exp.inflow	4.01e-03	4.38e-04	-1.02e-03	3.28e-04	-1.13e-05	3.85e-04
delta.exp.inflow	4.38e-04	4.66e-03	2.84e-04	-1.53e-03	1.54e-04	-9.95e-04
upper.gate	-1.02e-03	2.84e-04	2.66e-03	2.12e-04	1.37e-05	1.03e-04
delta.gate	3.28e-04	-1.53e-03	2.12e-04	4.69e-03	9.18e-05	-1.45e-03
mainstem.turbid	-1.13e-05	1.54e-04	1.37e-05	9.18e-05	3.43e-03	7.16e-04
delta.turbid	3.85e-04	-9.95e-04	1.03e-04	-1.45e-03	7.16e-04	4.47e-03

F S-PLUS parameter estimation program

This program can also be downloaded at
http://www.uidaho.edu/~newman/usfwsglm.html

```
 rr.glm.poisson() <- function(y, X = 0, myoffset = 1, k = 0, myfam = "poisson",
                    mylink = "log", chipps.n, ocean.n)
{
        #default offset = 1 has no effect given log link
        #always include intercept in this particular model

        #center and scale the covariates and add intercept
        xmeans <- apply(X, 2, mean)
        xsds <- sqrt(apply(X, 2, var))
        Xctr <- sweep(X, 2, xmeans, FUN = "-")
        Xctr <- sweep(Xctr, 2, xsds, FUN = "/")
        Xctr <- cbind(1, Xctr)
        p <- dim(Xctr)[[2]]

        #Initializations- using non-ridge glm
        temp <- glm(y ~ offset(log(myoffset)) + Xctr[, -1], family = poisson(
                link = log))
        mu <- fitted(temp)
        z <- log(mu) + (y - mu)/mu
        z <- z - log(myoffset)
        W <- diag(mu)
        eps <- 0.001
        err <- 1
        iter <- 0
        olddev <- 2

        #iterate between parameter estimation and "weight" calculations
        chipps.wt <- rep(1, chipps.n)
        while(err > eps) {
                iter <- iter + 1
                #parameter estimation
                param <- solve(qr((t(Xctr) %*% W %*% Xctr) + k * diag(p)), t(
                        Xctr) %*% W %*% cbind(z))

                #Update the Weight matrix and adjust dependent variable
                mu <- as.vector(myoffset * exp(Xctr %*% cbind(param)))
                z <- log(mu) + (y - mu)/mu
                z <- z - log(myoffset)
                chipps.disp <- sum((y[1:chipps.n] - mu[1:chipps.n])^2/mu[1:chipps.n])/
                               (chipps.n - p)
                ocean.disp <- sum((y[(chipps.n + 1):(chipps.n + ocean.n)] -
                              mu[( chipps.n + 1):(chipps.n + ocean.n)])^2/
                              mu[(chipps.n + 1):(chipps.n + ocean.n)])/(ocean.n - p)
                if(chipps.n > 0) {
                        ocean.wt <- rep(chipps.disp/ocean.disp, ocean.n)
                }
```

```
        else ocean.wt <- rep(1, ocean.n)
        W <- diag(mu * c(chipps.wt, ocean.wt))

        #Computations needed for convergence check
        #need to set zero value y's to small value for
        # the log step in deviation calculation
        temp <- y
        temp[y == 0] <- eps
        newdev <- 2 * sum(y * log(temp/mu) - (y - mu))
        err <- abs(newdev - olddev)/(olddev + eps)
        olddev <- newdev
    }
    resid <- y - mu
    pearson.chi.sq <- sum((y - mu)^2/mu)
    slopes <- param[-1]
    unstd.intercept <- param[1] - sum((slopes * xmeans)/xsds)
    unstd.slopes <- slopes/xsds
    unstd.param <- c(unstd.intercept, unstd.slopes)
    temp <- solve(t(Xctr) %*% W %*% Xctr + k * diag(p))
    var.z <- diag(c(rep(chipps.disp, chipps.n), rep(ocean.disp, ocean.n))/ mu)
    cov.beta <- temp %*% (t(Xctr) %*% W %*% var.z %*% t(W) %*% Xctr) %*% temp
    se.yhat <- sqrt(mu^2 * diag(Xctr %*% cov.beta %*% t(Xctr)))
    return(unstd.param,param,y,mu,resid,newdev,pearson.chi.sq,cov.beta,se.yhat,
    chipps.disp,ocean.disp)
}
```