**Preliminary late-fall Chinook salmon smolt outmigration analysis for SIT team, June 2016 meeting**

Prepared by Cyril Michel, UCSC/NMFS-SWFSC Santa Cruz Lab

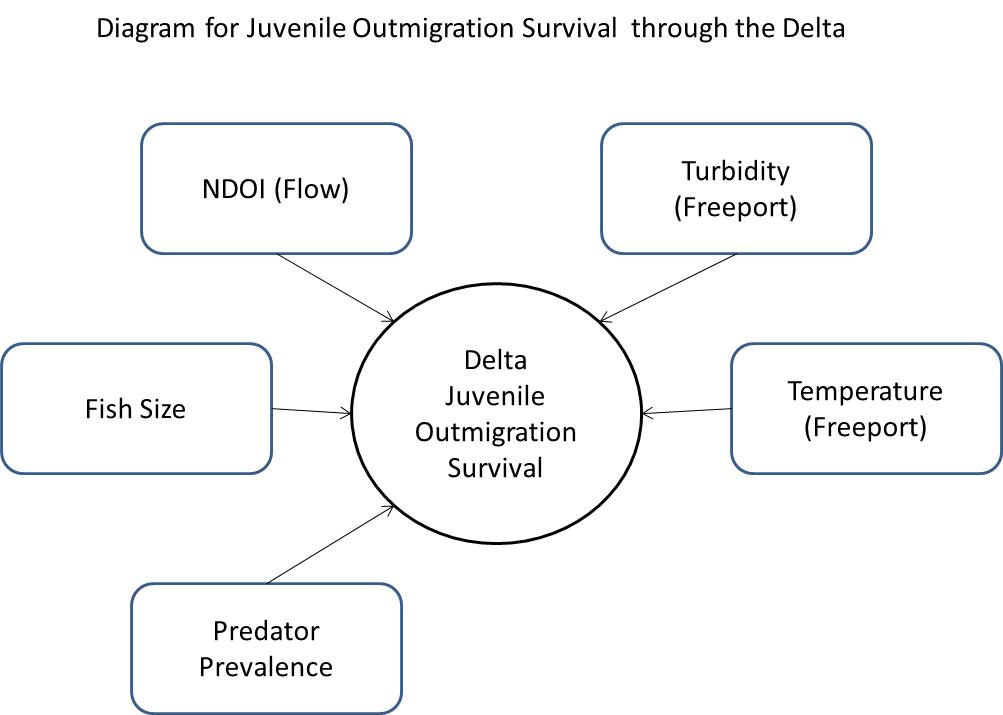
**Introduction**

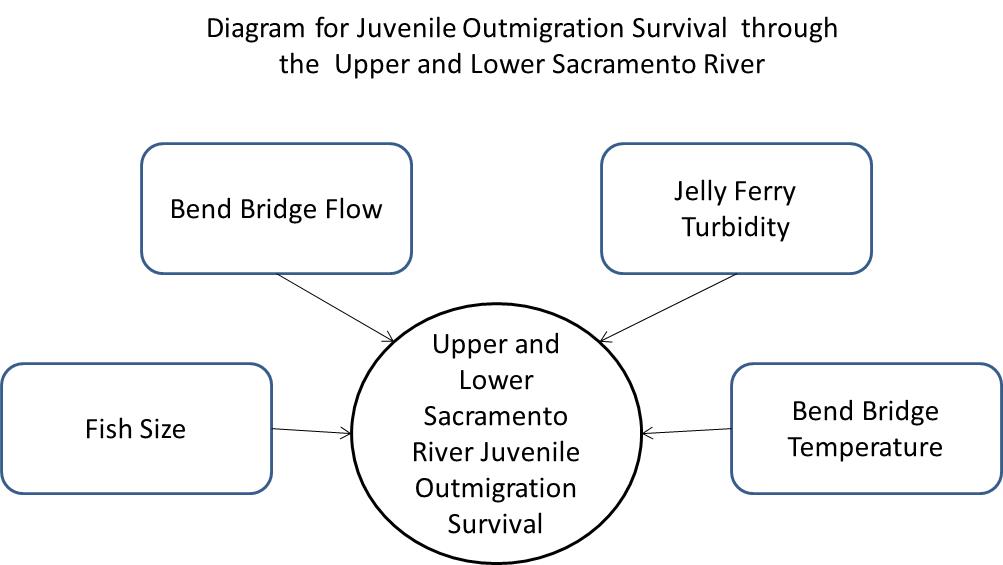
At the request of the SIT team, I have attempted to revisit late-fall Chinook salmon acoustic tagging data from winter 2007 to 2011. The intention of this new analysis is to elaborate on results published in Michel et al., 2015 (Chinoook salmon outmigration survival in wet and dry years in California's Central Valley. CJAFS. 72:1-11). These fish were tagged as part of CALFED funded grant to study survival dynamics of both late-fall Chinook and steelhead smolts in the Sacramento River.

Specifically, we would like to use mark-recapture modeling to investigate the potential influence of different environmental variables on Sacramento River region-specific survival of the smolts. I have done this through the RMark package in the R software program, which allows users to use Program Mark but through the interface of R. The following diagram shows the proposed environmental variables and their potential influence on different regions of the Sacramento River (fig.1).

**These analyses are preliminary in nature, and have not been peer-reviewed. As such, they should be used with caution, and should only be used for discussion purposes and guiding future analyses/investigations**. Furthermore, I am in the process of obtaining additional late-fall Chinook salmon acoustic tagging data from USFWS-Red Bluff. These additional data will likely improve our ability to detect relationships between environmental covariates and survival, and therefore final results are subject to change.

Figure 1.





**First steps/data massaging**

For the analysis, I had to first collect all environmental variables, then bring them into R, then organize them in such a way to merge all the environmental variables into one large dataframe. This was done fairly easily, but took a fair bit of data massaging once everything was loaded into R.

I also loaded all detections and tagging metadata into R. The original detections loaded are for all years, and monitor sites, and for all releases. However, for this analysis, I decided to not include fish that were released mid-river, and only kept fish released at Jelly's ferry or upstream, near the source of all these fish (Coleman Hatchery). This comprised removing approximately half of all tagged late fall, but I believe that including these fish would not be appropriate since we are trying to look at correlations between environmental variables and survival on a river-wide basis.

I also removed all detections from receivers that are not needed for this analysis. This will make all data summarizing steps quicker, since there are ~2million detections to begin with. I only kept detections at Jelly’s Ferry, Freeport, Chipps Island, Benicia, and the two lines at the Golden Gate Bridge. This effectively delineates the river, delta, Suisun bay, and SF bay, and the purpose of second Golden Gate line allows us to estimate survival to the first line, but the survival between the two lines is not meaningful.

In terms of study years, release strategies changed every year. In the first year, all fish were released in January of 2007, using a “trickle-out” method where ~13 fish were released per weekday for 4 consecutive weeks into Battle Creek. In the following 3 years, we released fish simultaneously from 3 different release sites (Jellys Ferry, Irvine Finch, Butte City) on one day in Dec and one day in Jan. Then in 2011, the last year, we released all fish from Jelly’s, one release day in Dec and one day in Jan. In the latter two years, USFWS service also tagged and released late-fall into Battle Creek along with production late-fall releases for one day in Dec and one day in Jan for both years. Those fish are not currently included in this analysis, but I have asked for permission to use them and should be able to for a future analysis. Finally, for all fish that were released at Battle Creek, I removed all detections of fish that were never detected at Jelly’s or further downstream since these fish likely died before passing Jelly’s Ferry, and therefore shouldn’t be included in the analysis.

Once all relevant detections had been subsetted, I used a crosstab analysis to summarize detections per location to a binary format, as used by program MARK. I then needed to find the environmental covariates experienced per fish per region (river and delta). The original analysis plan was find the means of the different river covariates over a 14 day period after release for the river section, and find the means of the different Delta covariates over a time window starting at (Release Date + Yearly median river transit time) and ending at [(Release Date + Yearly median river transit time) + Yearly median Delta transit time] for the Delta section. The 14 day window for the river was based off the median river transit time for all the tagged fish. However, river transit times did vary substantially per year, and therefore I decided it would make more sense to do the river covariate summarizing similarly to the Delta technique. In other words, I found the median river transit time per year, and then summarized the different covariates for the window starting at a fish’s release date, and ending at the fish’s (release date + Yearly median river transit time).

For this step, I first needed to estimate median transit time for the river and Delta. For the river, this consisted of finding the median value of all travel times of all the fish that were detected at Freeport, the “end” of the river region in this analysis, per year (Table 1).

Table 1

|  |  |
| --- | --- |
| **year** | **median\_river\_transit\_days** |
| 2007 | 11.34 |
| 2008 | 14.95 |
| 2009 | 14.91 |
| 2010 | 10.54 |
| 2011 | 12.98 |

For the Delta transit time, I found the median travel time of all fish that were detected both at Freeport (the “beginning” of the Delta region) to Chipps Island (the “end” of the Delta region) (Table 2).

Table 2

|  |  |
| --- | --- |
| **year** | **median\_Delta\_transit\_days** |
| 2007 | 6.68 |
| 2008 | 9.3 |
| 2009 | 7.41 |
| 2010 | 5.38 |
| 2011 | 5.5 |

A quick summary on each environmental variable:

-River\_flow: flow (CFS) collected from Bend Bridge gauge, summarized as daily median

-River\_temp: temperature (C) collected from Bend Bridge gauge, summarized as daily median

-River\_turb: turbidity (NTU) collected from Jelly’s Ferry gauge, summarized as daily median

-Delta\_flow: NDOI flow (CFS) collected from dayflow estimates, the OUT variable, daily estimate

-Delta\_temp: temperature (C) collected from the Rio Vista gauge, summarized as daily median

-Delta\_turb: an index created by multiplying suspended sediment to flow from the Freeport gauge, daily estimates

-Delta\_pred: predator prevalence (CPUE) from the IEP Mast report, yearly cumulative CPUE. This is the one variable that is summarized per year, rather than per day like all the others. I have included it at the request of the SIT team, but I don’t believe it’s entirely appropriate in this analysis.

Once mean environmental conditions experienced was calculated per fish, for the river and delta region separately, these were related to each fish’s encounter history, and the individual covariate of length was also appended to this dataset.

**CJS Modeling**

For the modeling portion of this analysis, I used the Cormack-Jolly-Seber model for live recaptures, using Program Mark through the interface of RMark package in program R. CJS models estimate both survival and detection probability.

The first modeling exercise is to better understand how the survival and detection probabilities change through time and space. I created all possible combinations of space and time effects for survival and detection probability (Table 3).

Table 3. Each row of this table represents one model, with the parameter structure for survival in column 1, and for detection probability in column 2. “~1” means the model has just one parameter, i.e. survival of detection probability is not allowed to vary over time ( year) or space (time). In other words, the ~1 ~1 model at the bottom of the table represents the null hypothesis that survival and detection probability are not a function of time or space. These models are ranked by AIC support, with the best model on top.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Survival** | **Detection Probability** | **Parameter Count** | **DeltaAICc** | **Weight** |
| Region \* Year | Region \* Year | 50 | 0.0 | 1.0 |
| Region | Region \* Year | 30 | 34.7 | 0.0 |
| Region \* Year | Region | 30 | 47.7 | 0.0 |
| ~1 | Region \* Year | 26 | 93.2 | 0.0 |
| Region | Region | 10 | 129.0 | 0.0 |
| Region \* Year | ~1 | 26 | 140.6 | 0.0 |
| ~1 | Region | 6 | 163.4 | 0.0 |
| Region | ~1 | 6 | 214.0 | 0.0 |
| ~1 | ~1 | 2 | 337.1 | 0.0 |

Table 3 indicates that there is strong support for region and year specific variation in both survival and detection probability. As a rule of thumb, a delta AIC of more than ~7 indicates substantially stronger support for the better model. Here, the (Region\*Year) model for both survival and detection probability is substantially better supported than any other model. Since we are only interested in modeling the effect of environmental variables on survival, I will set the parameter structure of detection probability of all upcoming models to “Region \* Year”. Additionally, the “~1” models has so little support that I will not include it in upcoming analyses. The survival estimates for the best model (Region \* Year) are provided in Table 4 below. They indicate similar patterns as seen in Michel et al, 2015: relatively low river and San Francisco Bay survival compared to other regions, with the exception of 2011 when a 2+ fold increase in survival was seen in the river region compared to previous years which results in the highest total outmigration survival rates of the five year study.

Table 4. Survival, standard error, lower and upper confidence interval estimates for the Survival (Region \* Year) Detection (Region \* Year) model.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **year** | **region** | **Survival** | **SE** | **95% lower confidence level** | **95% upper confidence level** |
| 2007 | river | 0.17 | 0.06 | 0.09 | 0.32 |
| 2007 | delta | 0.64 | 0.22 | 0.22 | 0.92 |
| 2007 | suisun | 0.64 | 0.21 | 0.23 | 0.91 |
| 2007 | SFBay | 0.43 | 0.19 | 0.14 | 0.77 |
| 2008 | river | 0.39 | 0.07 | 0.26 | 0.53 |
| 2008 | delta | 0.49 | 0.1 | 0.3 | 0.69 |
| 2008 | suisun | 0.67 | 0.11 | 0.43 | 0.84 |
| 2008 | SFBay | 0.46 | 0.14 | 0.22 | 0.72 |
| 2009 | river | 0.36 | 0.05 | 0.27 | 0.46 |
| 2009 | delta | 0.43 | 0.09 | 0.27 | 0.6 |
| 2009 | suisun | 1 | 0 | 0 | 1 |
| 2009 | SFBay | 0.26 | 0.11 | 0.1 | 0.52 |
| 2010 | river | 0.37 | 0.05 | 0.27 | 0.47 |
| 2010 | delta | 0.58 | 0.09 | 0.4 | 0.75 |
| 2010 | suisun | 0.55 | 0.11 | 0.34 | 0.75 |
| 2010 | SFBay | 0.25 | 0.12 | 0.08 | 0.55 |
| 2011 | river | 0.7 | 0.09 | 0.49 | 0.84 |
| 2011 | delta | 0.66 | 0.09 | 0.47 | 0.82 |
| 2011 | suisun | 0.82 | 0.05 | 0.71 | 0.89 |
| 2011 | SFBay | 0.39 | 0.05 | 0.29 | 0.5 |

The next step of the analysis is to one-by-one incorporate different environmental covariates to a base model as you would an individual covariate, allowing us to compare the model support between each covariate model as an indication of support for that covariate. As a base model, I will use the “region” survival model. I have chosen this model and not the better supported “region \* year” model because the purpose of this modeling exercise is partly to identify what covariates drive year to year differences in survival, and therefore we hope that the covariate parameter will explain the year to year fluctuations in lieu of the year parameters.

Table 5. Each row of this table represents one model, with all models sharing a detection probability parameter structure of (Region \* Year). Most models contain one environmental variable, with the exception of the “Region” model, which will serve as our base model on which we will add covariates, and our “full” model, with allows survival to vary per region and per year. Each environmental covariate is only allowed to act upon the region from which it was collected, e.g. for the Region + river\_flow model, all reaches are allowed to have different survival estimates, and furthermore, the river region has a linear relationship with river\_flow covariate. Finally, individual fish length is added as a covariate, both as an additive model and a multiplicative model. An additive model means fish length can have a linear relationship with region-specific survival, but this relationship doesn’t change from region to region; a multiplicative model mean fish length can have a linear relationship with region-specific survival, and the slope of this relationship can change from region to region.

|  |  |  |  |
| --- | --- | --- | --- |
| **Survival** | **Parameter Count** | **DeltaAICc** | **Weight** |
| Region \* Year | 50 | 0.0 | 0.5 |
| Region + river\_flow | 31 | 0.3 | 0.5 |
| Region + delta\_pred | 31 | 9.1 | 0.0 |
| Region + delta\_flow | 31 | 14.5 | 0.0 |
| Region + delta\_temp | 31 | 23.7 | 0.0 |
| Region + river\_temp | 31 | 24.2 | 0.0 |
| Region + river\_turb | 31 | 34.0 | 0.0 |
| Region | 30 | 34.7 | 0.0 |
| Region + delta\_turb | 31 | 35.7 | 0.0 |
| Region + length | 31 | 36.9 | 0.0 |
| Region \* length | 35 | 42.0 | 0.0 |

Table 4 indicates a strong relationship between flows in the river section and survival. In fact, the river\_flow model is as well supported as the “full” model. In other words, including river\_flow in the survival model accounts for approximately the same amount of variation in survival estimates as allowing survival estimates to vary year to year. These two models are by far the two best supported models, with the model allowing for predator CPUE to have a linear relationship with survival coming in next best. However, as mentioned before, this result is likely not meaningful because the predator CPUE are on a yearly time step while all the other variables are on a daily time step. Furthermore, looking at the beta parameter estimates of both the river\_flow and delta\_pred covariates, the river\_flow beta parameter is positive, indicating that increases in flow are correlated with increases in survival. However, the delta\_pred beta parameter is also positive, suggesting somewhat counterintuitively that an increase in predators in the delta correlate with increases in survival. Finally, there is marginal support for the delta\_temp and river\_temp models over the Region (base) model off which they were constructed. All other models have similar or lower AIC values as the base model, and therefore are not supported. In other words, this preliminary modeling exercise has not found support for turbidity being an important environmental variable, nor has it found support for a relationship between survival and fish length.

The next step of this analysis was to investigate if combining the influence of the same environmental variable on their respective different regions in one model may further increase the support over the current set of models. For example, instead of river\_flow and delta\_flow being separate models, combine them so that one model allows river survival to vary in relationship with river\_flows and delta survival to vary in relationship with delta\_flows (see Table 6).

Table 6. This table has the same model outputs as table 5, with 3 additional models added highlighted in yellow. These models are models that allow for river conditions to have an effect on river survival, and for the same condition but in the delta to have an effect on delta survival.

|  |  |  |  |
| --- | --- | --- | --- |
| **Survival** | **Parameter Count** | **DeltaAICc** | **Weight** |
| Region \* Year | 50 | 0.0 | 0.5 |
| Region + river\_flow | 31 | 0.3 | 0.4 |
| Region + river\_flow + delta\_flow | 32 | 2.4 | 0.1 |
| Region + delta\_pred | 31 | 9.1 | 0.0 |
| Region + delta\_flow | 31 | 14.5 | 0.0 |
| Region + river\_temp + delta\_temp | 32 | 19.1 | 0.0 |
| Region + delta\_temp | 31 | 23.7 | 0.0 |
| Region + river\_temp | 31 | 24.2 | 0.0 |
| Region + river\_turb | 31 | 34.0 | 0.0 |
| Region | 30 | 34.7 | 0.0 |
| Region + river\_turb + delta\_turb | 32 | 35.3 | 0.0 |
| Region + delta\_turb | 31 | 35.7 | 0.0 |
| Region + length | 31 | 36.9 | 0.0 |
| Region \* length | 35 | 42.0 | 0.0 |

As table 6 indicates, the three additional combined covariate models are not necessarily better supported than the single covariate models of the same kind. Of particular interest, the flow combined covariate model is less well supported than the simple river\_flow model, further cementing this model as the best supported environmental model in this preliminary modeling effort.

This next section is exploratory, and is a work in progress. The next logical step in this preliminary analysis is to better understand the relationship between river\_flow and river survival, in particular the slope of the relationship. The current river\_flow model allows for a simple linear relationship between river flow and river survival. But perhaps the relationship between river flow and river survival is better explained by a linear relationship including a squared term, which could better fit the relationship if it happens to be asymptotic, for example. Finally, perhaps the relationship is even more complex, and a quadratic relationship is required to capture it. I have created a basic model of these two in Table 7.

Table 7. These are top 4 models of the Table 6 after additional river\_flow survival models have been added.

|  |  |  |  |
| --- | --- | --- | --- |
| **Survival** | **npar** | **DeltaAICc** | **weight** |
| Region + river\_flow2 | 31 | 0.0 | 0.5 |
| Region \* Year | 50 | 1.8 | 0.2 |
| Region + river\_flow | 31 | 2.2 | 0.2 |
| Region + river\_flow + river\_flow2 | 32 | 2.5 | 0.1 |

Table 7 seems to indicate that the relationship between river\_flow and survival may be better explained by a squared relationship, although only marginally better supported than the “full” model and the simple linear river\_flow model.

We can then create a simulation plot showing the relationship of the two best environmental models with survival. We do this by simulating multiple regularly spaced values ranging from the minimum to the maximum flow level experienced by these smolts. We than use the flow beta parameter estimate to predict what the survival estimate would be given each flow level. This allows us to visualize the relationship, potentially looking for “threshold” flow levels above which survival increases substantially (a stated objective of the SIT team). The two following figures show this relationship for the two best fit environmental models: (Region + river\_flow; fig. 2) and (Region + river\_flow2; fig.3).

Figure 2. The simulated relationship between the flow and survival in the river section using the beta parameter estimates for the flow covariate, dotted lines represent 95% confidence intervals.

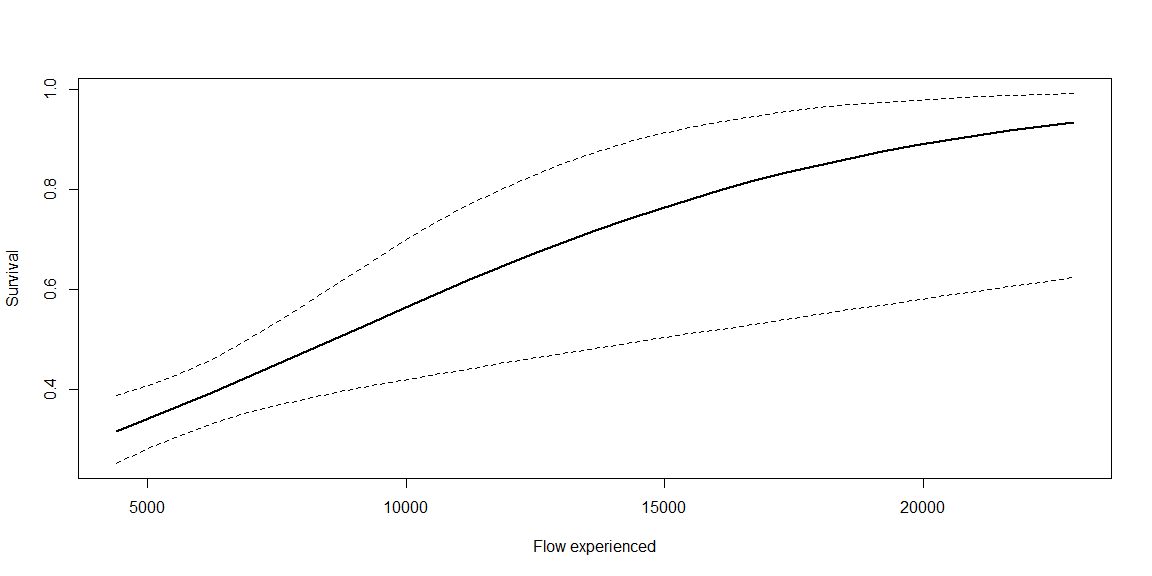
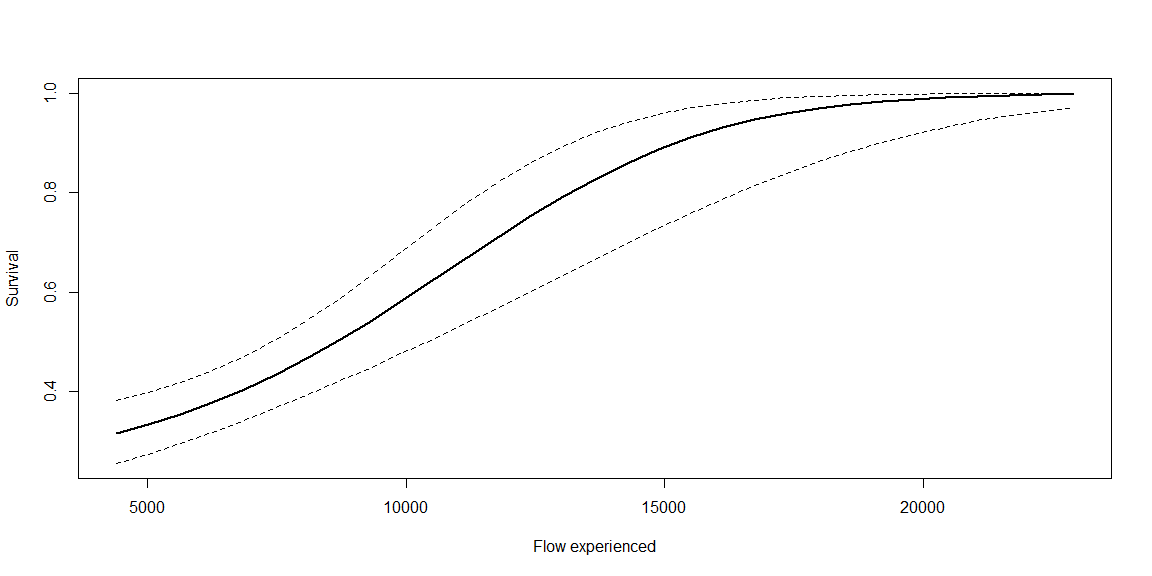


Figure 3. The simulated relationship between the flow and survival in the river section using the beta parameter estimates for the flow2 covariate, dotted lines represent 95% confidence intervals.



Figures 2 and 3 show a fairly strong relationship between flow and survival, where survival can vary hypothetically vary from 0.35 to almost 1 under varying flow conditions. We don’t see a strong “threshold” or point of inflection signature, but it does seem from figure 3 that between 10000 and 15000 cfs, there are potentially substantial gains in survival.

Future directions

This analysis is still very preliminary, but some promising future directions have presented themselves. They include:

-Adding more tagging Late-fall Chinook to the analysis, in particular the USFWS fish

-Adding the CALFED tagged fish that were released mid-river as part of this analysis. They might not be appropriate for inclusion into the river region analyses, but could be useful for Delta analyses.

-Adding additional environmental variables of interest

-Further testing hypothetical relationships between flow and survival