CVCS Escapement Modeling

Arthur Barros

2024-06-11

## Introduction

This document aims at provide an overview of the modeling methods used to estimate the annual escapement, the number of adults returning to spawn, of Central Valley Chinook Salmon (CVCS). Through this document we will review the basic concepts of mark-recapture methods, more complex models like the Cormack-Jolly-Seber, and how to write scripts that can calculate escapement estimates. This review will heavily rely on utilizing the R scripting, and assumes the reader has a base understanding of the language, how to write and implement R scripts. I also assume the reader has a baseline understanding of statistical concepts such as linear regression and probability.

This document is part of an effort to better translate and understand the current CVCS escapement model utilized by California Department of Fish and Wildlife (CDFW) and other groups in the escapeMR package. This model and R package and application was written by the contractor Trent McDonald in the 2020, and while it has proven very useful, it acts as a sort of “black-box model” in which users input data and receive outputs without understanding of how it works. The goal of this documentation is to provide interpretation of that model so that scientists can make better decisions about its implementation, and make improvements when applicable.

## An Introduction to Mark-Recapture Models

Mark and recapture is a common method for ecologists to estimate the size of a population of animals without needing to count each individual. Typically an initial ‘marking’ period is done, in which a sample of the population is captured and marked in some way so that those individuals can be recognized in the future. Then the marked individuals are returned to the greater population and allowed time to mix back in. Later, another sampling period will be conducted in which a sample of the population is collected again, and the number of marked individuals in this sample are enumerated. Because the proportion of marked individuals in the second sample should be proportional to the number of total marked individuals in the population, this ratio can be used to estimate the size of the total population.

In regards to the CVCS escapement mark-recapture surveys are conducted throughout the Central Valley on CVCS carcasses in what is known as a “carcass survey”. Periodically, field crews will conduct surveys of Chinook salmon spawning grounds, looking for adult carcasses, which they mark with disk tags and return to the river, usually returning within a week to see if the carcass is still in the system, and mark new carcasses that have arrived. The field methods are a little more complex than that summary, and will be expanded upon in further sections.

### *The Lincoln-Petersen estimate*

The classic mark-recapture estimator is known as the **Lincoln-Petersen** ([Petersen 1896](#ref-Petersen1896); [Lincoln 1930](#ref-Lincoln1930)) and utilizes the equation:

Where is the number of individuals marked and released during the initial marking period; is the number of marked individuals recaptured in the later recapture period; is the number of individuals captured in the recapture period; and is the estimate of population abundance.

In the diagram below (Figure 1), 7 fish from the population are captured and marked () during an initial capture and marking period. Then during a follow up recapture period, 7 fish are again captured (), and of those 3 have markings (). This leaves us with an estimate of a total population of roughly 16 fish.

|  |
| --- |
| Figure1: A conceptual diagram of the classic Lincoln-Petersen mark-recapture method. |

There are several important assumptions of the Lincoln-Petersen method:

1. The population is closed so that is constant during the study period. No deaths, no births, no immigration or emigration.
2. The probability of capture is equal across all individuals.
3. Fish do not lose their markings between the two sampling periods.
4. All marked fish recaptured during the second sampling period are accurately enumerated and recorded.

The Lincoln-Petersen method was expanded upon upon by Zoe Schnabel ([1938](#ref-Schnabel1938)) with the **Schnabel estimator**, which expands the Lincoln-Petersen estimator to allow for multiple periods of capture and recapture, instead of just two.

This allows for marking to occur during multiple sampling periods . One primary advantage of this method is multiple sampling periods allow for the detection of violation of the above assumptions. Regression of the proportion of marked animals to previously marked animals () will be straight if assumptions are met, and curved if the assumptions are violated.

If we examine our the annual carcass surveys performed across the Central Valley for returning Chinook salmon, we can easily identify ways in which the system violates the assumptions of the Lincoln-Petersen method:

* **Populations are open:** throughout the survey period adult Chinook arrive in the survey area, spawn, and die. At the same time spawned out carcasses that may or may not have been marked and counted are removed from the system either by scavengers, decay, or being flushed downstream.
* **Probability of capture is not homogeneous:** probability of”capture” for individual Chinook carcasses is really the probability it is spotted by survey field crews. This can be dependent on a number of factors including carcass size, carcass sex, water flow rates, and turbidity.

The above violations of the Lincoln-Petersen estimates could cause large biases in abundance estimates, and suggests that using this closed population model is ill advised.

### *The Cormack-Jolly-Seber models*

To deal with open populations and heterogeneous capture probabilities we can utilize a Cormack-Jolly-Seber (CJS) model ([Cormack 1964](#ref-Cormack1964); [Jolly 1965](#ref-Jolly1965); [Seber 1965](#ref-Seber1965)), which can deal with population changes during the survey period, as well as differences in capture probability among individuals. The CJS method does not directly provide estimates of abundance, instead it utilizes multiple sampling events during the survey period and allows us to calculate two important parameters:

* : the probability carcass “survives” (remains in the system) from period to .
* : the probability carcass is captured during period .

With these probabilities the capture history of a fish can be estimated for each sampling period as:

We will generate estimates for using the CJS Maximum Likelihood Estimation methods later, but for now its important to know that it can be utilized to estimate the population size at period using the “Horvitz-Thompson estimator” ([Horvitz and Thompson 1952](#ref-Horvitz1952); [Nichols 2008](#ref-Nichols2005)):

$$
\hat{N\_j}=\sum\_{i=1}^{n}\frac{h\_{ji}}{\hat{p}\_{ji}} \space \text{ (eq. 3)}
$$

Where is the 0 or 1 capture indicator for carcass at period , and is the number of all carcasses observed over all survey periods. Basically it’s saying the abundance of carcasses at time is equal to the sum of all the carcasses captured during that period divided by their capture probability.

Typically the total population estimate is calculated by taking the average value of , however there are problems with this method when many individuals are entering and leaving the population during the survey. This can be improved by using a *superpopulation* modification ([Ryan, Cooper, and Tauer 2013](#ref-Ryan2013)) to estimate total escapement as:

$$
\hat{N}\_{escapement}=\hat{N}\_2\frac{ln(\tilde{\phi}\_1)}{\tilde{\phi}\_1-1}+B\_2^\*+B\_3^\*+...+B\_{S-2}^\* \space \text{ (eq. 4)}
$$

Where is the number of “births” (new adults entering the system, spawning, and dying) between and is:

$$
B\_j^\* = \tilde{B}\_j\frac{ln(\tilde{\phi}\_j)}{\tilde{\phi}\_1-1} \space \text{ (eq. 5)}
$$

is the total births for period :

$$
\tilde{B}\_j=\hat{N}\_{j+1} - \tilde{\phi}[\hat{N}\_j-(n\_j-R\_j)]\space \text{ (eq. 6)}
$$

is the number of carcasses released with disctags during period , is the number of carcasses captured during period , and is the estimate of abundance during period derived from the Horvitz-Thompson estimator in equation 3.

## The Math Behind the CVCS Escapement Model

#### **The Data**

For the CVCS carcass surveys, carcasses are captured over occasions, usually weekly sampling events, marked with disctags, and deposited back into the system for future recapture. The CJS model used for CVCS carcass surveys incorporates three different data inputs to estimate and : capture histories, “chops” records, and covariate data.

**Capture histories** are the records of capture and recapture for a given fish and usually look similar to:

Here each record is for a tagged carcass and the subsequent sampling periods . A value of 0 indicates the carcass was not captured, a value of 1 indicates the carcass was captured, and a value of 2 indicates the carcass was captured and “chopped” (beheaded) and removed from the system.

**Chops** is the record of the number of carcasses removed from the system upon first capture. This is done when the carcass is deteriorated beyond identification or measurement, and wont last long enough for further recaptures. The data can look like:

With just one row of total chopped counts for each sampling period.

**Covariate** **data** is the record of relevant covariates for each captured carcass or each survey period. For CVCS surveys, the covariates currently used are the sex and the length of the carcass. In the future we hope to incorporate environmental variables such as temperature and flow. A covariate table usually looks like:

With the covariates recorded for each carcass marked with a disctag, and used to help determine different values of and for each carcass.

#### Estimating Probabilities of Capture and Survival

While the above equations used to estimate abundance using carcass survey data seem relatively straight forward, they are heavily dependent on estimates of and , which are much more complex to get.

Let’s say we have the following capture history for a fish , across four different sampling periods:

We can assign a probability of observing that capture history as:

Note that in the above, is the probability that the carcass is not captured at period . Now we can do the same with the rest of the example capture histories that we saw:

In the above note the 2 at the end of the last capture history representing a chopped event. Chopped carcasses will be dealt with later. We can use those equations to estimate the total likelihood of observing all the capture histories we saw as:

Or as the product of all of the capture history probabilities:

To make the computing of this probability easier we can log-transform our likelihoods so that:

This is known as the “log-likelihood” of the probability. Using this simple example above, we have a way to estimate the probability of a set of given capture histories if we have the capture () and survival () probabilities associated with each capture history.

Of course, we don’t know these probabilities, and they will likely vary across different covariates. For CVCS work, we utilize covariates of sex and length to try and estimate capture and survival probabilities. For our example we will consider a CJS model in which capture probability is a function of length, ans survival probability is a function of sex.

These functions can best be represented as logistic models:

Where is the sex of a given carcass, and it’s length. To ease in estimating the coefficients and we can turn the above into linear models:

You can see above that if we can estimate the coefficients of those two equations, for a given sex or length we can estimate the corresponding probability of survival or capture. To estimate the coefficients of , , , and , we can use a method know as **maximum log-likelihood estimation**. To do this we first need to be able to estimate the log-likelihood of our model on a given set of coefficients and capture histories, basically: what is the log-likelihood of the capture histories we observed given a set of coefficients?

If we refer back to equation 7, we can remember that the log-likelihood of all the observed capture histories is equal to the sum of the log-likelihood of each individual capture history. The log–likelihood of the capture history for a given carcass can be given as the sum of its capture and survival probabilities from its first to last encounter, plus : the probability it is not captured after its last release. This individual capture history log-likelihood can be represented as:

Where is 1 if the carcass was observed during , and 0 otherwise. can be represented as:

The first part of this equation represents the probability the given carcass is removed from the system immediately after its last observation. The second part of equation 11 is the product of all probabilities that carcass remained in the system but wasn’t observed during all subsequent observation periods.

Now we have an equation (eq. 10) that provides us with the log-likelihood of observing a given capture history with specific survival and capture probabilities. We have also shown that values of and can be linear regressed across the coefficients , , , and (equations 8 and 9). So now we can plug in values for the coefficients of equations 8 and 9 to give us capture and survival probabilities, which we can put into equation 10 to give us the log-likelihood of observing the capture history of that carcass using those coefficients. We can repeat this process until we find the coefficients of equations 8 and 9 that provide us with the highest log-likelihood.

Of course, repeating the above steps to find the highest log-likelihood for even just one carcass history would take some time to do manually, and we need to find the coefficients that will provide the highest likelihood across all of our capture histories. Fortunately, we can use an optimization algorithm to automate this process, and find the coefficients that produce the maximum likelihood estimation for us. We’ll discuss the possible algorithms that can be used in the following sections.

Once the optimization algorithm provides us with the coefficients that give us the highest likelihood of observing our capture history data with the given covariates, we can use it to calculate the capture probability for each carcass () given it’s length and sex. Those probabilities can then be utilized in equation 3 to estimate total escapement.

## Coding the Modified CJS Model

### Reviewing the escapeMR package

### Coding the Functions Ourselves

pro\_capsur<-function(beta,i,j,p){  
 #purpose: evaluate probability of capture and survival for each animal i  
 #global vars: lengths\_matrix,sex\_matrix,nx,nan,ns  
 cap\_beta<-beta[1:nx]  
 surv\_beta<-beta[(nx+1):np]  
 zp<-exp(cap\_beta[1]\*1+cap\_beta[2]\*cap\_X[i,j])  
 p.hat<-zp/(1+zp)  
 zs<-exp(surv\_beta[1]\*1+surv\_beta[2]\*surv\_X[i,j])  
 s.hat<-zs/(1+zs)  
 est\_list<-list('p.hat'=p.hat,'s.hat'=s.hat)  
 return(est\_list)  
}

location<-function(nan,ns,ch){  
 #purpose: compute first and last capture for each animal  
 first <- rep(0, nan)  
 last <- rep(0, nan)  
 for(i in 1:nan){  
 findic=TRUE  
 for(j in 1:ns){  
 if(ch[i,j]>=1){  
 if(findic==T && (j<ns)){  
 first[i]=(j+1)  
 findic=FALSE  
 }  
 last[i]=j  
 }  
 }  
 }  
 est\_list<-list('first'=first,'last'=last)  
 return(est\_list)  
}

## Testing the Model

## References

Cormack, R. M. 1964. “Estimates of Survival from the Sighting of Marked Animals.” *Biometrika* 51 (3): 429–38.

Horvitz, D. G., and D. J. Thompson. 1952. “A Generalization of Sampling Without Replacement From a Finite Universe.” *Journal of the American Statistical Association* 44 (260): 663–85. <http://www.ncbi.nlm.nih.gov/pubmed/16243365>.

Jolly, G. M. 1965. “Explicit Estimates from Capture-Recapture Data with Both Death and Immigration- Stochastic Model.” *Biometrika* 52 (1): 225–47.

Lincoln, Frederick C. 1930. “Calculating Waterfowl Abundance on the Basis of Banding Returns.” Vol. 118. US Department of Agriculture. <https://arxiv.org/abs/arXiv:1011.1669v3>.

Nichols, J. D. 2008. *Modern Open-Population Capture-Recapture Models*. Edited by Steven C. Amstrup, Trent L. McDonald, and Bryan F. J. Manly. Princeton Univ. Press.

Petersen, C. G. J. 1896. “The yearly immigration of young plaice in the Limfjord from the German sea.” *Report of the Danish Biological Station* 6: 5–84.

Ryan, Cooper, and Tauer. 2013. “CENTRAL VALLEY CHINOOK SALMON IN-RIVER ESCAPEMENT MONITORING PLAN.” *Paper Knowledge . Toward a Media History of Documents*, 12–26.

Schnabel, Zoe E. 1938. “The Estimation of the Total Fish Population of a Lake.” *American Mathematics Monthly* 45: 348–52.

Seber, G A F. 1965. “A Note on the Multiple-Recapture Census Published.” *Biometrika* 52 (1): 249–59.