METHODS

As previously described, my dataset includes natural-origin salmonid spawner abundance estimates for 68 populations of Chinook and coho salmon and steelhead trout in Oregon. Data covers the years 1980 to 2022 and includes the survey method employed for each observation. A total of 31 different survey methods were employed in collecting this data. These methods will be the focus of this project, as it is my goal to assess the level of precision and directional bias of each (a full list of survey methods is included as an appendix).

This analysis will model each salmonid species separately, leading to three data subsets respectively containing 18 Chinook populations, 27 coho populations, and 23 steelhead populations. With data already sorted by species, each observation contains four crucial pieces of information: 1) an estimate of natural-origin spawner abundance (denoted mathematically from here on as *y*), 2) the population of interest (denoted with the subscript *p*), 3) the survey method used (denoted with the subscript *s*), and 4) the year of observation (denoted with the subscript *t*).

Being that the goal of this analysis is to isolate variance and bias in different survey methods, it is vital to distinguish variability in observations due to survey method from variation in the natural state of salmonid populations. Multivariate autoregressive state space (MARSS) models are well suited for this purpose, as they are able to mathematically distinguish observation error from process/state error and assign separate variance parameters to each. A simplified MARSS model is described below:

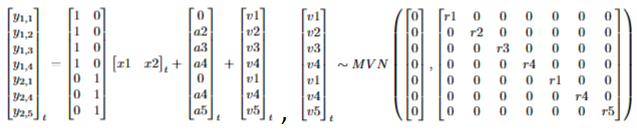
Observation equation: *yt* = **Zt***xt* + *at* + *vt*, where *vt* ~ MVN(0, **Rt**)

Process/state equation: *xt* = **Bt***xt-1* + *ut* + *wt*, where *wt* ~ MVN(0, **Qt**)

Critical to these models is the matrix **Zt**, which assigns each of *n* observations in the vector *yt* to one of *m* states in the vector *xt*. **Zt** may be specified such that every observation in *yt* is representative of its own underlying state (in which case *n* = *m* and **Zt** is an identity matrix), such that every observation in *yt* is representative of one underlying state (in which case *m* = 1 and **Zt** is a column vector of 1s), or specified to assign observations to any number of states in between (crucially *m* must always be less than or equal to *n*).

Applying the MARSS model structure to the analysis described above involves determining how best to structure spawner abundance data (*yp,s,t*) and specify model parameters to isolate precision and bias in salmonid survey methods. Particular attention is paid to the inputs and parameterization of the observation equation. *yt* is arranged such that each row relates to observations of population *p* by survey method *s*. **Zt** is specified such that each population as identified in the data will correspond to a distinct underlying natural state. By specifying each population as an underlying state, the model will be able to assess variation in the natural state of the population as separate from variation in counts arising as a result of the survey method employed. The vector *at* is specified such that each survey method is assigned a separate *a* value irrespective of population with the intention of capturing each method’s directional bias. One method will be selected as a reference against which other methods will be measured, meaning that for this method *a* is set equal to zero. The matrix **Rt** is specified in a similar manner, with a unique *r* term for each method placed in the corresponding places along **Rt**‘s diagonal. Again, this is irrespective of population in order to capture the variation inherent in each survey method.

Below is an example of the observation equation written out in matrix form. In this example, two populations of salmonid have been surveyed with a total of five survey methods. As indicated by the *y* term subscripts, population one has been surveyed with methods one through four and population two with methods one, four, and five. **Zt** is specified with two columns relating to the two populations which here are considered underlying states (*x1* and *x2*). Survey method one has been selected as the reference method against which other methods’ directional bias will be measured. Therefore, in the vector *at*, those rows which correspond to an observation taken with method one are set equal to zero, while all other methods are represented by an *a* parameter which repeats as often as the method appears. Finally, **Rt** is specified so that each survey method is represented by an *r* parameter which again repeats as often as the method appears, in order to capture the variance inherent to each.



The process/state equation is more straightforward. It is constructed to be as parsimonious as possible, in that it in this analysis it essentially acts as a receptacle for the natural variance in annual salmonid spawner returns. **Bt** is specified as an identity matrix, meaning the whole equation becomes a random walk. Whether the vector *ut* is set to zero or allowed to estimate a trend in each population’s growth will depend on model fit. Finally, the terms along the diagonal of **Qt** will be set equal to each other in initial model specifications, although the inclusion of other specifications (unequal population variances and equal variance and covariance) will also depend on model fit. Below is an example of the process/state equation written out in matrix form.



The various specifications of *ut* and **Qt** will be evaluated for best fit based on AICc, as it is better suited for balancing fit with the many parameters MARSS models must estimate than AIC.

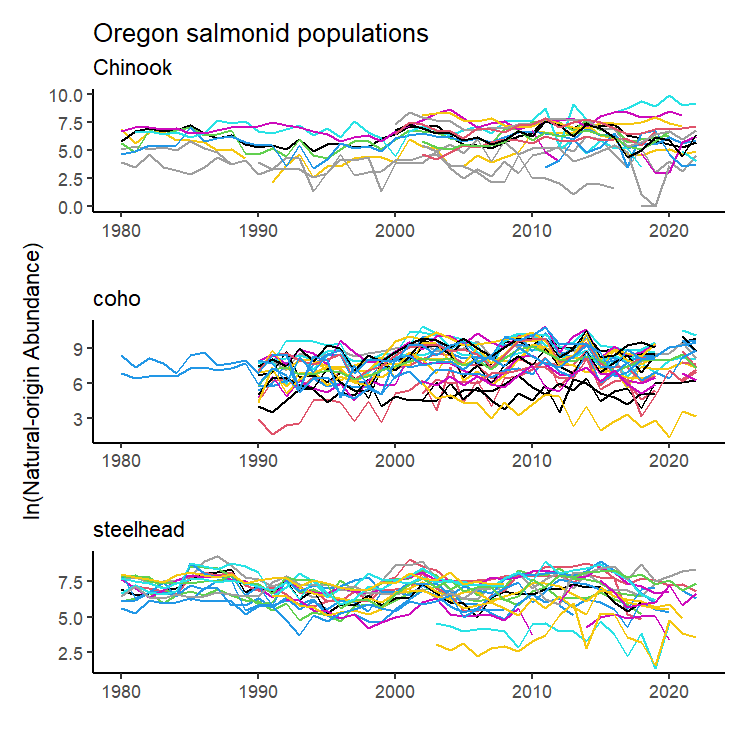
Attached is R code which structures survey data and specifies MARSS model parameters according to the procedure described above and the data necessary for the code to run. Also included in this document is an appendix describing the 31 survey methods included in my dataset, as well as the project proposal and draft introduction.

APPENDIX – SURVEY METHODS



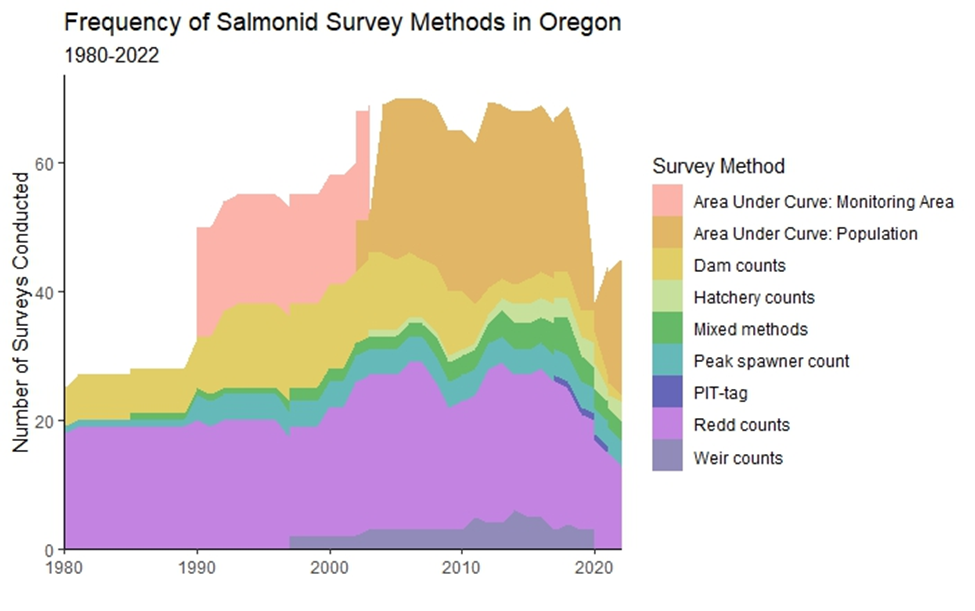
PROJECT PROPOSAL

My research project seeks to examine the efficiency of various salmonid survey methods employed by the Oregon Department of Fish and Wildlife. My intention is to evaluate 31 survey methods to determine their precision and potentially identify any directional bias.

Data includes annual observations of “natural-origin salmon and steelhead abundance” or NOSA, covering 69 populations of Chinook and coho salmon and steelhead trout throughout Oregon from 1980 to 2022. Crucially, each observation also includes a variable indicating the method by which the abundance survey was conducted. Data has been collected through 31 unique survey methods. The data contains 2,924 individual observations, of which 848 are missing NOSA values which almost entirely relate to years in which no survey was conducted (15 observations confoundingly include a survey method along with a null value). Figure 1 illustrates changes in abundance of individual salmonid populations throughout the study period.

*Figure 1.*

The frequency with which survey methods are employed is not uniform across the study period. Figure two illustrates groups of survey methods and the frequency of their employment from 1980 to 2022. The nine groups displayed in the figure are bins containing each of the 31 survey methods described in the data. While some methods such as (redd counts) have been widely used throughout the study period, other methods have seen drastic changes in usage.

I intend to analyze this data by fitting MARSS models and comparing estimates of variation related to different survey methods. Specification of these models will require careful consideration to capture the precision of each survey method, as the individual time series (each related to a particular population of salmonids) are typically composed of observations drawn from multiple survey methods. I *think* this will likely take the form of careful construction of the MARSS model’s **R** matrix to isolate the behaviors of various sampling methods.

*Figure 2.*

DRAFT INTRODUCTION

The conservation of salmon and trout for environmental, tribal, commercial, and recreational use is of tremendous importance to state-level fish and wildlife agencies in the pacific northwest. Oregon alone has, since 1980, spent over $9 billion in the effort to protect and restore populations of steelhead trout and salmon. To determine the value of investments like these, policymakers and fisheries managers need accurate assessments of the populations in question. To this end various survey methods may be employed, with each method accompanied by some cost and presumably characterized by an inherent level of precision.

This study seeks to measure the levels of precision accompanying various survey methods by comparing observations from 69 populations of Chinook and coho salmon and steelhead trout throughout Oregon, taken from 1980 to 2022. Additionally, this work seeks to determine whether directional bias exists in some survey methods, although any estimates of bias will likely be relative to other survey methods, as “true” censuses of the underlying salmonid populations do not exist. The intention of this research is to map the value of information, such that managers can better determine achievable levels of precision given some budget, or inversely the cost of some minimum level of precision in their estimates.