Stream points (chapter 2) methods detail

In chapter 2, I am predicting how many hatchery strays would go to index streams in SE AK which have not been surveyed, i.e., there is no information on how many hatchery salmon were at those streams between 2008­–2019. This requires compiling data for my significant model covariates based on the final model from chapter 1. These were 1) conspecific abundance (Cons\_Abundance), 2) the number of fish released within 40km (WMA\_Releases\_by\_Yr), and 3) the coefficient of variation of flow (CV\_flow). Pink salmon abundance (Pink\_Abundance) was not included in the final model, but it had been in an earlier version, so as a result I had tailored a Pink\_Abundance dataset for chapter 2. The Pink\_Abundance dataset was the most complete and comprehensive of the covariate datasets for chapter 2 (contains stream name, number, AWC code, lat, long, etc, for +600 total streams), so I used it as the main dataset to join the other datasets to. Specific methods details are described:

In *pink\_salmon\_counts.csv* (see “PinkDens&Abundance” folder from chapter 1), which was shared with me by Andy Piston and gives the data reported in the 2020 Piston Heinl pink salmon stock status report for SEAK, there are 706 streams with pink salmon count data. This pink salmon count data is associated with a stream number, not an AWC code (see *Pink\_Methods\_Detail.docx* for chapter 1 in “PinkDens&Abundance” folder for more information. I linked these streams to the correct AWC # using the *stream\_num\_to\_AWC.csv* file (“PinkDens&Abundance” folder) and then linked them to the stream data downloaded from Mapping Data Files for the Anadromous Waters Catalog for all streams documented to have chum salmon usage (*All\_SEAK\_CH\_Strms.csv*, n=2038). This gave me the name, AWC #, stream number, subregion, and pink salmon count data for 706 SEAK streams containing pink and chum salmon. I double-checked all 706 streams for correct stream number assignment and name-matching and deleted those with missing location data, AWC #s, or those whose names did not match between datasets the AWC data and *stream\_num\_to\_AWC.*csv data. This left with me a final total of 662 streams for chapter 2 predicting (including the original 59 streams from chapter 1). This final dataset is called *Chp2\_StreamPoints.csv.*

* Then I joined the other covariate datasets (described below) to this dataset to create the *Chp2\_Master\_Dataset.csv*. See *create\_master\_data.R* for more information

Pink salmon abundance methods detail

Using a dplyr ‘join’ in R, I linked the stream (n = 706) pink salmon count data from *pink\_salmon\_counts.csv* (“PinkDens&Abundance” folder; chapter 1) to *Chp2\_StreamPoints.csv* using the stream numbers. *Chp2\_StreamPoints.csv* is the datafile containing stream names, numbers, AWC numbers, and location data for the 662 streams I am analyzing in chapter 2. Methods for creating and tailoring *Chp2\_StreamPoints.csv* are described in the section above.

I randomly double-checked approximately 10% of the 662 total pink salmon data stream assignments for accuracy. The final dataset containing this pink salmon abundance data for chapter 2 is called *Pink\_Abundance\_Chp2.csv*. For more information on pink salmon data and methods, see *Pink\_Methods\_Detail.docx* in the “PinkDens&Abundance” folder from chapter 1.

Conspecific abundance methods detail

Andy Piston emailed chum salmon abundance data on 8/9/2021 for the 87 streams in SEAK for which chum salmon abundance is monitored. This includes the original 64 streams from chapter 1 and is the same data as *Chum\_salmon\_counts.csv* from chapter 1 (only with 23 additional streams), which is also included in the 2020 Piston and Heinl chum salmon stock status report. The file is called *Cons\_Abundance\_Chp2.csv.*

**Important note regarding chapter 2 conspecific abundance:**

In chapter 1, Cons\_Abundance in the model was the estimated count of chum salmon in a stream multiplied by 1-pHOS for the stream, that way the hatchery contribution to the escapement would be factored out. I don’t have any pHOS data on streams for chapter 2 (hence why I am predicting their attractiveness), so the Cons\_Abundance data for chapter 2 streams is just the chum salmon escapement count not multiplied by anything.

I included the chum salmon escapement count for streams within the chapter 2 master dataset that were also included in chapter 1. The Cons\_Abundance data for these OG streams are the Cons\_Abundance data NOT multiplied by 1-pHOS, so you can compare model output for streams which did have data in chapter 1 based on the methods difference for Cons\_Abundance between chapter 1 and chapter 2.

I also manually entered data for 4 streams from “Chum\_counts\_by\_stream.xlsx” in chapter 1 which were interpolated (“Chum\_counts\_by\_stream.xlsx” contains the conspecific abundance estimates NOT multiplied by 1-pHOS). That is why there would be 40 more rows of Cons\_Abundance data in Chp2\_Master\_dataset compared to what I output from my R script where I originally created the master dataset. These streams were

1. Ushk
2. Camp Coogan
3. Ketchikan Creek
4. Staney Creek

CV\_flow methods detail

Same methods as in chapter 1. See “Hydro\_methods\_detail.docx” in “Stream\_hydro\_Flow” folder

WMA\_Releases within 40km methods detail

Similar to WMA\_Releases methods for chapter 1, I determined all release sites within 40km of chapter 2 streams using the *Measure* tool on google maps (<https://www.google.com/maps/d/u/0/edit?mid=1-qPyZA6jEJIZHxxO7G1mbGfBuVmgUuvx&ll=53.96564415045508%2C-136.3721285&z=5>) and transferred this into R. This file is called “Releases\_within40km\_chp2.csv”. I assigned a weighted moving average (WMA) of the number of fish released 2, 3, and 4 years prior to the theoretical year of sampling based on expected ages of returning adults (3, 4, and 5yo, which would have been released from release sites 1 year after birth, or 2-4 years prior to sampling)

* That is, for each stream’s release site(s), I assigned a vector of the study years (2008-2019, not including 2012 and 2016). To each of these years, I assigned another vector of each year minus 2, 3, and 4. Like this:

|  |  |  |  |
| --- | --- | --- | --- |
| Stream | ReleaseSite | Year of “sampling” (had the stream been sampled) | Release year (2 years prior for 3yo, 3 years prior for 4yo, etc) |
| Sawmill Creek | AMALGA HARBOR 111-50 | 2008 | 2006 |
|  |  |  | 2005 |
|  |  |  | 2004 |
|  |  | 2009 | 2007 |
|  |  |  | 2006 |
|  |  |  | 2005 |
|  |  | 2010 | 2008 |
|  |  | Etc.. through 2019 | 2009 |
|  | BOAT HARBOR 115-10 | 2008 | 2006 |
|  |  |  | 2005 |
|  |  |  | 2004 |
|  |  | 2009 | 2007 |
|  |  |  | 2006 |

* Then for each release site + release year row in the above table, I assigned the corresponding release site data (E.g., AMALGA HARBOR 2006 = 34.645) and multiplied these numbers by the ***proportion*** of 3, 4, and 5yos in the data
  + ***Proportions 🡪*** I summed up the total # of 3, 4, and 5yos (based on markIDs which indicate brood year) in the entire 2008-2019 hatchery wild dataset and determined their proportions relative to one another. There were 3.8% 3yos, 61.2% 4yos, and 35.0% 5yos. AMALGA HARBOR 111-50 2006 release would have been multiplied by 0.038 for 3yos for 2008
* I summed proportions of releases by fish age for each release site to get a weighted moving average of the number of fish released 2-4 years earlier for each release site. Finally, I summed up this WMA value for all release sites within 40km of a stream, that way I would have the total released within 40km of a stream, not divided up by release site.
* See “Releases40km\_chp2.R” for more details and to see the code

Also note that in chapter 1, my weighted moving average was calculated based on the proportion of 3–6-year-old fish a given stream. For chapter 2, I was using the total proportion of each age across time to calculate a WMA, and you may notice that I only did this for fish ages 3–5, not 3–6. This is because relative to the total, there were very few 6-year-olds, so I didn’t bother factoring that into my calculation. Below is the table of the number of each age of fish

|  |  |
| --- | --- |
| Age | Total number in HW Data from 2008–2019 |
| 3 | 320 |
| 4 | 5105 |
| 5 | 2915 |
| 6 | 148 #not worth including this fraction |
| 7 | 4 #definitely not worth including |

Other notes about data

The chapter 1 streams are included in the chapter 2 dataset and have the same data as in chapter 1 for pink salmon abundance, mean flow, and CV flow. Their data values differ slightly for

1. Conspecific abundance because I do not have 1-pHOS values to multiply the conspecific abundance estimates by for chapter 2 streams as I did in chapter 1
2. WMA\_Releases\_by\_Yr because I do not have an empirical age distribution for chapter 2 streams either (no hatchery salmon sampled to determine ages from)

For chapter 1 streams included in the chapter 2 dataset, I did methods for Cons\_Abundance and WMA\_Releases\_by\_Yr the same way I did for chapter 2 streams (no 1-pHOS multiplier nor stream-specific age distribution) so that I can compare model output for these streams which have a known response and predictions based on chapter 1 methods. Hopefully the predictions between the datasets don’t differ greatly!