Hatchery Indicators Methods

1. All hatchery releases made by WDFW in Region 5 were queried from FishBooks to identify distinct combinations of brood year, species, run, stock, and facility, and adipose clip type.
2. All populations located in Region 5 were queried from the “stock” table in SPi. Some populations required editing their population names because they were either missing or incorrect in order to match the NOAA naming conventions.
3. All SPi indicator data was queried for Region 5.
4. An attempt was made to join the results of 1 & 2 to identify which NOAA designated populations were used in spawn events where unclipped fish were spawned. This join was attempted by manipulating normalizing species names and run types between the basins, and then calculating string distances between the two data sources based on species and run and comparing the “location” portion of each NOAA population name with the facility and stock fields in the FishBooks query. The result of this attempt was saved as “DRAFT\_R5\_hatchery\_stock\_NOAA\_pop\_LUT.csv”
5. This LUT was manually inspected and edited to be correct. The corrected version was saved as “R5\_hatchery\_stock\_NOAA\_pop\_LUT.csv”
6. The LUT was then loaded and joined to a fishbooks query of all unique brood codes to assign the correct NOAA population used for each brood code that utilized wild broodstock in the Region 5 (only for salmon and steelhead). To determine which brood codes used wild fish, code was used to assign the origin of production as wild or hatchery using a mixture of Adipose and CWT status and basin-specific rules. Brood codes with either W x W parents or W x H parents were combined (resulting in HxH and mixed brood codes)
7. PNOB, pHOB were calculated for each brood code, stock, and year
8. Hatchery releases were queried for Region 5 and unique sites identified.
9. An LUT was created to identify anadromous and non-anadromous release sites to filter releases to anadromous waters
10. NOAA population polygons were loaded from the NOAA geodatabase and spatial joins were used to assign releases to all populations (across multiple species/runs) those release sites belonged to. Several release sites required corrections to their lat/long with rationale listed in the code
11. Once release sites were mapped to populations, all release events were joined to the release sites first by release location, species, and run. For any releases that remained unjoined a second attempt was made to join them by release location and species only. This ensured the pHOS of, for example, hatchery summer steelhead spawning in the Elochoman River, would be counted toward the only steelhead population in that basin (winter run) if there was no designated wild steelhead population in that basin with the same run type as the summer run hatchery releases.
12. Finally, the “R5\_hatchery\_stock\_NOAA\_pop\_LUT.csv” was joined to the releases such that a NOAA population was assigned as the broodstock source if the production included wild parents.
13. All releases were summed by brood year, species, and NOAA population, and weighted averages (across all releases) were calculated for the proportion of natural origin brood (pNOB) and the proportion of hatchery origin brood (pHOB) using those values for the associated brood codes and weighting by the number of each brood code released in each NOAA population in each brood year for each species.
14. Only in cases where wild parents were spawned and the progeny were released within the same wild populations boundaries was the production considered integrated (when only hatchery fish were spawned or when progeny of wild fish were released outside their wild population boundaries, production was considered segregated and pNI was not calculated).
15. Finally, SPi indicator data (pHOS, NOSA, TSA) was joined by release site NOAA population and brood year with the releases.
16. We then calculated the remainder of the indicator values pNI as pHOS/(pHOS + pNOB).
17. We developed an R ShinyApp to display and retrieve data by NOAA population.
18. Complete code to reproduce this analysis is located here: <https://github.com/tbuehrens/R5hatchery_indicators>. Note that the code makes use of passwords, usernames, and database addresses that must be stored in an R environment file which is not included as part of the GitHub Repository. To run the code valid access credentials must be stored in your R environment file with identical names.