Modifications to the Chum Genetic and Environmental Management Model (ChumGEM), a run reconstruction model

2019 Annual Report prepared by

Lingbo Li[1], Brittany Jenewein[1], and Pieter Van Will[2]
Fisheries and Oceans Canada,
1 Annacis Island, Unit 3 – 100 Annacis Parkway
Delta, BC V3M 6A2
2 Port Hardy Field Office, 8585 Wollason Street
PO Box 10, BC V0N 2P0

on behalf of the ChumGEM Working Group, a subset of the Pacific Salmon Treaty Chum Technical Committee

for

Angus Mackay
Fund Manager
Pacific Salmon Commission
600-1155 Robson Street
Vancouver, B.C. V6E 1B5

Table of Contents

List of Tables

List of Figures

1.0 Introduction

Sustainable fisheries management is challenging when and where multiple populations of varying productivity are harvested together. To reduce the likelihood of management errors that may cause serious conservation harm to depleted populations and economic harm to the fisheries targeting abundant populations, salmon managers need to understand the expected abundance in each fishery. The Chum Genetic and Environmental Management Model (ChumGEM), a run reconstruction model, was developed between 2013 and 2015 to assist in post-season run reconstruction analyses for Chum Salmon fisheries in southern British Columbia and Puget Sound, Washington. The model uses a genetic data set, historic catch data, and escapement estimates to provide estimates of run-specific parameters such as abundance and arrival timing. It also provides daily estimates of exploitation rates and escapement by genetic unit (GU).

Substantial modifications are needed to make the ChumGEM model a useful tool for fisheries management. Testing of the model was conducted in 2017, which concluded that the model should still be considered a work in progress pending further data verification, model checks, and sensitivity analysis. Furthermore, changes to the model structure are likely required. The objective of this project is to modify and improve the ChumGEM model so that it can be used as a fisheries management tool.

2.0 Major modifications and corrections to ChumGEM

Many modifications and corrections have been made to the original ChumGEM model. Substantial efforts have been made to figure out how the model worked and what modifications are required. The following major changes are highlighted based on their large impacts on the model and great efforts needed.

2.1 Identifying Chum Salmon GUs

A total of ten fall GUs were identified based on consistent Genetic Stock Identification (GSI) surveys. GSI surveys have been consistently conducted during early September and early

November in Canadian Management Areas 12 and 13 in the Johnstone Strait since 2013 and in the Strait of Juan de Fuca (JdF; Area 20 in Canadian waters and Area 5 in U.S. waters) since 2016. Due to the survey timing, ten fall GUs were identified and included in the model while another six GUs of summer and winter runs were excluded (Table 1). Placeholders for these GUs remain in the model code and could be included in future analysis if surveys are expanded to occur during the migration timing of these GUs.

An index of proportion of each GU to the whole Chum population was developed to show the relative abundance across stocks based on consistent GSI data in Johnstone Strait and JdF. P_{sya}, the average proportion of fall GU s in year y and GSI sampling area a, was estimated as:

$$P_{sya} = \frac{\sum_{i=1}^{k} N_i \cdot P_i}{\sum_{i=1}^{k} N_i} \qquad (1)$$

Where N_i represents number of fish in sample i; P_i represents proportion of stock s in sample i; k represents total number of GSI samples in year y and area a. The optimum sample size (N_i) is 200 fish but this was not always achieved (in some cases < 50 fish were sampled). This method can minimize biases introduced by small sample sizes within a year and area.

This index shows Canadian GUs, particularly Fraser (FR) followed by Strait of Georgia East (SOGE) and Strait of Georgia West (SOGW), dominate the Chum population in Canadian Management Area 12 (Figure 1), whereas Hood Canal (HC) occasionally contributes substantially in JdF (Figure 2). This was not consistent with the original priors that showed mean run size of HC was three times that of FR. Priors of run size were therefore re-estimated for each GU (Figure 3).

2.2 Defining diversions of Chum Salmon GUs

The index P_{sya} also demonstrates three diversions for each GU because each GU was present in the following sampling areas: Canadian Management Area 12 (Figure 1) and Canadian and U.S.

pathways of JdF (Figure 2). Each GU migrates along a northern (N) path through Johnstone Strait (JS) to the Strait of Georgia and also along two southern (S) paths through both the Canadian and U.S. waters of JdF. The average diversion rates between N and two S paths were first based on experts' estimation. Originally, there was only one southern pathway for each stock; a Canadian pathway for Canadian stocks and a U.S. pathway for U.S. stocks. Based on our analyses of the average proportion of each GU in the sampling areas, two southern diversions were recently developed for each stock. This means that each GU now has the potential to migrate through both Canadian and U.S. waters, rather than only one or the other. Generally, however, Canadian fish migrate mostly through CDN waters and U.S. fish migrate mostly through U.S. waters.

It was noticed that northern diversion fish may split into two around Texada Island after migrating through JS into the Strait of Georgia. On the west side of the island is Canadian Management Area 14 and on the other side is Canadian Management Areas 15 and 16. Catch in Area 14 is substantially higher than in Areas 15 and 16. Based on expert knowledge, the pathways around Texada Island were corrected: SOGW, WCVI, and all U.S. GUs use only Area 14, SOGE uses only Areas 15 and 16, and FR evenly splits on either side of the island.

2.3 Modifying pathways

In terminal areas, large discrepancies have been found between the original model and data. The model included some terminal areas where there are no reported escapements, but also sometimes excluded some terminal areas where large escapements occurred. Making the model terminal areas consistent with data was the first step.

The original model split each diversion evenly into all the terminal areas of each GU, which created three major problems: 1) This greatly increased model complexity and model inefficiency. The original model has 202 pathways in total. For example, Puget Sound South (PSS) originally had eight terminal areas with two diversions, thus generating 16 pathways. Some terminal areas are so close together (e.g. about 10-15 km) that a fish would take a half-day or less to migrate from one area to another, so the usefulness of including these small areas was questioned given the migration time is less than the time step for fish movement currently used in the model. This problem may become more serious when fish swim around or hold instead of returning directly to their home stream. The number of total pathways will be further amplified as there are now three diversions as opposed to the original model, which had between 1 and 3 diversions for each GU. 2) How to split each GU into different pathways poses great challenges. Within a single GU, many more fish return to some terminal areas than others (e.g. HC); because of this, the even-splitting method may introduce large biases in both catch and escapement estimation for each pathway. Unfortunately, as the escapement has large inter-annual variability among areas, it is very hard to establish a mechanism to accurately split the populations. 3) The escapement likelihood sums escapement by GU instead of by GU and terminal area. Constrained by the data resolution, there is no reason to modify the escapement likelihood to the stock and terminal area level (See section 2.7 Correcting the likelihood function). Therefore, it is not useful to split fish into different terminal areas.

A completely new pathway structure, about 35 pathways in total, is being built and is expected to greatly improve the model. We will remove the original terminal areas and identify the single terminal area that has the longest route as the only terminal area for each GU. This terminal area will represent the total escapement for each GU. Therefore, three diversions lead to three pathways from the beginning of model area to the terminal area for each GU. An exception is the GU of Puget Sound North (PSN); likely because the northern and southern terminal areas are geographically separated, the northern and southern fish are also genetically distinct (Maureen Small unpublished data). Therefore, two terminal areas were developed for PSN, making six pathways. All these modifications greatly reduce model complexity and increase efficiency, while fully meeting our management goals.

2.4 Probability of fish movement from one area to another

The method used to determine the probability of fish movement from one area to another was improved. Originally, no matter how large an area is, all fish moved to the next area the day after arriving in the preceding area. However, the model assumes fish have a normal distribution of migration rate with a mean of 37.7 km/day and standard deviation of 6 km/day. Some areas are greater than 60 km long; therefore, it is highly unlikely a fish can pass through these large areas within one day. The corrected migration is more realistic. For example, in Canadian Management Area 16 with a distance of 30 km, in the next day, 36% of fish will stay in Area 16, 63% will move to the next Management Area, and 1% will go to a further area. This also improves the total number of fish and hereafter the catch estimation in each area.

2.5 Correcting the exploitation rate estimation

The exploitation rate estimation was corrected by constraining it to be less than 1. Time- and area-specific exploitation rates were estimated by dividing the observed catch by the estimated total number of fish at a given area and a given day:

$$U'_{at} = \frac{C_{at}}{\sum_s \sum_x N_{sxat}} \tag{2}$$

Where *s* represents stock group (GU); *x* represents pathway; *a* represents management area; and *t* represents day. C_{at} represents the observed catch at area *a* and day *t* and N_{sxat} represents the estimated number of fish in GU *s* of pathway *x* at area *a* and day *t*.

Based on expert knowledge, the exploitation rates of Chum Salmon in southern BC and Puget Sound are expected to be much less than one. However, it is possible for the model to predict the total number of fish close to or larger than the observed catch, generating an exploitation rate close to or greater than one. The model originally replaced all estimated exploitation rates higher than 0.95 with uniform 0.95:

$$U_{at} = \begin{cases} U'_{at} & U'_{at} \leq 0.95 \\ 0.95 & U'_{at} > 0.95 \end{cases} \tag{3}$$

These high exploitation rates (>0.95) were corrected by adjusting them to new values less than one:

$$U_{at} = \begin{cases} U'_{at} & U'_{at} \leq 0.95 \\ 1 - \frac{1}{400\,(U'_{at} - 0.9)} & U'_{at} > 0.95 \end{cases} \qquad (4)$$

As shown in Figure 4, the corrected function (Equation 4) is a significant improvement because higher exploitation rates still have higher adjusted rates while being constrained to less than one. The cutoff of 0.95 was chosen based on expert knowledge and it is consistent with a Sockeye run reconstruction by Branch and Hilborn (2010).

2.6 Validating catch and escapement data

As outlined in the project objectives, all catch data were validated and included in the model as the amount of fish taken from the population. Originally only commercial data were used as catch data in the model. Other available catch has now been added to the model, including data from the First Nations Food, Social, and Ceremonial fisheries, test fisheries, and recreational fisheries.

Canadian escapement data were infilled based on historic visual surveys. Estimates derived using peak counts or Area-Under-the-Curve (AUC) from multiple surveys tend to be the main estimation methodology. Even though there are issues with the quality of some of the escapement data (e.g. variation in observers and variability in the levels of coverage), they are sufficient to understand overall trends in the abundance over time. The periodicity of escapement monitoring in the Inner South Coast (ISC) region has resulted in significant gaps in the dataset. To reconstruct the escapement for all Chum-producing streams in a given Conservation Unit (CU), the gaps in the escapement time series need to be infilled. Data compiled from the New Salmon Escapement Database System (NuSEDS) was grouped by CU prior to applying the infill technique. The geometric mean of each stream’s escapement across years was calculated and assigned as the stream’s average escapement. The total average escapement for the CU in a given year was the sum of all stream average escapements within that CU and year. The proportion of monitored escapement covered in a given year was determined by summing the average escapements of all streams monitored in a given year divided by the total average escapement for all streams in that area. The reconstructed escapement of a given CU and year was then calculated as the sum of observed escapements for that year and CU divided by the proportion of monitored escapement for that year and CU. This method of reconstructing the escapement was applied to the seven ISC Chum Salmon CUs (four GUs: JS, SOGW, SOGE, JdF). As the infilled escapement data are used in the model, GU-specific variances in observation errors are not considered at this stage.

2.7 Correcting the likelihood function

Likelihood function has been greatly improved. The purpose of this function is to compare different parameter values and select the ones that result in the best fit of the model to the data. As the model includes prior distributions and likelihood functions, model outputs are the modes of the Bayesian posterior distribution. The function has been modified to minimize the sum of the negative log-likelihood of three factors:

$$-\ln L = -\sum_{i=1}^{3} \ln L_i \qquad (5)$$

The corrected likelihoods now include catch (L_1), escapement (L_2), and stock proportion (L_3) with each as the sum of differences between the estimated and the observed values. Catch likelihood is estimated as the sum of differences between the estimated total catch and observed total catch for each area and each day:

$$-\ln L_1 = \sum_a \sum_t (\ln \hat{C}_{at} - \ln C_{at})^2 \qquad (6)$$

Where C_{at} represents catch at a given area *a* and a given day *t*, with $\hat{C}$ indicating estimated catch and *C* indicating observed catch. It is important to understand that the catch likelihood (L_1) presents a penalty only when the estimated exploitation rate is higher than 0.95 and needs to be adjusted (See section 2.5 Correcting the exploitation rate estimation). When the estimated exploitation rate is lower than 0.95, the estimated total catch always equals the observed catch in a given area and day as shown in the following equations:

$$\hat{C}_{sat} = U_{at} \sum_x N_{sxat} \quad \text{(See } U_{at} \text{ in Equations 2 and 4)} \qquad (7)$$

$$\hat{C}_{at} = \sum_s C_{sat} \tag{8}$$

For the escapement likelihood (L_2), we considered minimizing the sum of differences in escapements by GU and Management Area between the observed and the estimated (Equation 9). However, due to the limitation of data resolution at the Management Area level, we decided to keep the original escapement likelihood as the sum of differences in total escapement in all areas by GU (Equation 10).

$$-\ln L_2 = \sum_s \sum_a (\ln \hat{E}_{sa} - \ln E_{sa})^2 \tag{9}$$

$$-\ln L_2 = \sum_s (\ln \hat{E}_s - \ln E_s)^2 \tag{10}$$

The prior likelihood function was removed from the overall likelihood function. The priors include mean run size, standard deviation of run size, mean arrival date to the beginning of the model, Area C10, and standard deviation of arrival date. The prior likelihood function forced the model estimation to be as close to the priors of the mean values as possible, overusing the prior and greatly underweighting the observed data including catch, GSI and escapement. This explains why the priors had a large influence on the model outputs as reported in the previous sensitivity analyses. Excluding priors in the likelihood function helps to better estimate the run size and arrival parameters and other parameters such as exploitation rates and escapements.

Catchability likelihood was also removed from the model. Based on catchability of similar fish species, it is believed unreasonable to assume constant catchability for Chum Salmon as in the original model. Michielsens and Cave (2018) examined catch, GSI, and hydroacoustic data of Fraser Sockeye Salmon, which showed that catchability varies greatly among stocks and days. Semi-pelagic Walleye pollock were also observed to change their horizontal and vertical distribution in response to environmental changes such as bottom depth, light levels, water temperature, sediment size and water currents (Kotwicki et al. 2013); Consequently, gear catchability varies in space and time. Furthermore, due to a lack of data, it is impossible to estimate catchability for each Chum GU during their return migration in different environmental conditions. This constant-catchability likelihood structure was therefore removed, including catchability estimation based on test fishery data and the likelihood of minimizing the differences between the daily catchability and average catchability.

3.0 Objectives re-evaluated as low priority

1. Change the model structure so that multiple years are fit simultaneously. Although sharing information across years could improve model fit significantly, this was put as low priority. With many problems identified in the original model, we decided it was more important to complete other improvements and modifications to ensure the single-year model works well before incorporating the multiple-year structure.

2. We also put low priority to incorporating age-specific data for the same reason as above.

3. We considered infill procedures for escapement and GSI data. However, due to large gaps in the GSI sampling, we decided infilling the GSI data would introduce too much bias to the model and opted not to continue with the procedure. The escapement infill procedures have been reviewed and we determined no changes are required at this time.

4.0 Next steps for improving the model

There is a strong need to weight the three likelihood functions that contribute to the overall model likelihood function. The magnitude of escapement varies greatly; some GUs range from more than half a million fish (e.g. FR) to a couple of hundreds of fish (e.g. JdF). Without weighting, the FR GU will dominate the escapement likelihood and several small GUs will be completely ignored in the model. By changing the magnitude of escapement for different GUs to put them on the same scale, we can change the weighting of each GU in the escapement likelihood and make each GU contribute more equally to the model. Furthermore, the magnitude of the likelihood also varies substantially among the three likelihoods (catch, escapement and stock composition). Further weighting is therefore needed across the three likelihoods to ensure they are more equally represented in the overall likelihood function.

The diversion rates at JS and JDF are expected to be further improved. They are currently set as fixed ratios for each GU. We plan to use the model to optimize the diversion rates. If it does not work well, we will run sensitivity analyses to assess the influence of different diversion rates.

Migration of fish within the model needs to be further modified. The probability of fish moving from one area to another has been improved but there are many other related issues. For example, fish migration starts at a fixed point of each area every day without accounting for the locations in the previous day. The accumulated errors in migration can be large. There are also uncertainties in migration rates. Based on previous tagging data, the migration rate is set for each GU as a normal distribution with a mean of 37.7 km/day and standard deviation of 6 km/day. Recent test fishery data in JS and at Albion (in the Fraser River) shows an average of 35 km/day. Sensitivity analyses will be conducted on the mean and standard deviation of migration rates.

5.0 Future studies that may further improve the model

Catchability is believed to vary substantially among GUs spatially and temporally. However, no data are available for Chum Salmon in the southern British Columbia and Puget Sound. This may have large negative influences on our stock-specific estimations of catch and hereafter escapement. There is a strong need to investigate how catchability varies among GUs and environmental factors.

Based on the design of the pathways in the model, fish from GUs that would not pass the terminal area of a neighbouring GU are not expected to contribute to catch in that terminal area. However, this may not be true; a small proportion of other GUs may also be present in some terminal areas because fish do not always travel directly to their home stream. For example, in terminal Sockeye Salmon fisheries in Alaska, Sockeye from neighbouring river systems have been reported as contributing significantly to the catch (Menard and Miller 1997; Habicht et al. 2007; Dann et al. 2009). Cunningham et al. (2018) presents the first attempt to account for this issue in the run reconstruction model. Based on expert knowledge and some GSI data, we believe that this could be a concern for Chum Salmon in some of our study areas. Due to a lack of consistent quantitative GSI data in terminal areas, we did not consider accounting for this issue in the current model. Further GSI surveys and investigation of this issue are strongly recommended.

6.0 Acknowledgements

We would like to thank the Pacific Salmon Commission's Boundary Restoration and Enhancement Fund for their financial support to this project. We also greatly appreciate insightful discussions had with Chum Salmon Technical Committee, particularly Bill Patton and Joe Tadey, and with the Pacific Salmon Commission, particularly Dr. Catherine Michielsens, Merran Hague, and Steve Latham.

7.0 References

Branch, T.A., and Hilborn, R. 2010. A general model for reconstructing salmon runs. Canadian Journal of Fisheries and Aquatic Sciences 67(5): 886–904.

Cunningham, C. J., Branch, T. A., Dann, T. H., Smith, M., Seeb, J. E., Seeb, L. W., & Hilborn, R. 2017. A general model for salmon run reconstruction that accounts for interception and differences in availability to harvest. Canadian Journal of Fisheries and Aquatic Sciences 75: 439– 451.

Dann, T.H., Habicht, C., Jasper, J.R., Hoyt, H.A., Barclay, A.W., Templin, W.D., Baker, T.T., West, F.W., and Fair, L.F. 2009. Genetic stock composition of the commercial harvest of sockeye salmon in Bristol Bay, Alaska, 2006–2008. Fishery Manuscript Series 09-06. Alaska Department of Fish and Game, Anchorage, Alaska.

Habicht, C., Seeb, L.W., and Seeb, J.E. 2007. Genetic and ecological divergence defines population structure of sockeye salmon populations returning to Bristol Bay, Alaska, and provides a tool for admixture analysis. Trans. Am. Fish. Soc. 136(1): 82–94.

Kotwicki S., De Robertis A., Ianelli J. N., Punt A. E., Horne J. K. 2013. Combining bottom trawl and acoustic data to model acoustic dead zone correction and bottom trawl efficiency parameters for semipelagic species. Canadian Journal of Fisheries and Aquatic Sciences 70: 208–219.

Menard, J., and Miller, J.C. 1997. Report to the Alaska Board of Fisheries on the stock composition of sockeye salmon catches within east side Bristol Bay fishing districts, 1983–1995. Regional Information Rep. 2A97-31. Alaska Department of Fish and Game, Division of Commercial Fisheries Management and Development, Anchorage, Alaska.

Michielsens, C.G., and Cave, J.D. 2019. In-season assessment and management of salmon stocks using a Bayesian time–density model. Canadian Journal of Fisheries and Aquatic Sciences 76(7): 1073-1085

Table 1. Genetic Units (GUs) represented in the ChumGEM run reconstruction model. Ten fall GUs (bold) are included in the model. The rest remain in the model only as placeholders due to a lack of genetic stock identification data.

GU	GU name	Run season
FR	**Fraser River**	**Fall**
JS	**Johnstone Strait**	**Fall**
SOGE	**Strait of Georgia East**	**Fall**
SOGW	**Strait of Georgia West**	**Fall**
WCVI	**West Coast Vancouver Island**	**Fall**
HC	**Hood Canal**	**Fall**
JDF	**Juan de Fuca**	**Fall**
PSC	**Puget Sound Central**	**Fall**
PSN	**Puget Sound North**	**Fall**
PSS	**Puget Sound South**	**Fall**
HC	Hood Canal	Summer
JDF	Juan de Fuca	Summer
PSS	Puget Sound South	Summer
PSC	Puget Sound Central	Summer
PSC	Puget Sound Central	Winter
PSS	Puget Sound South	Winter

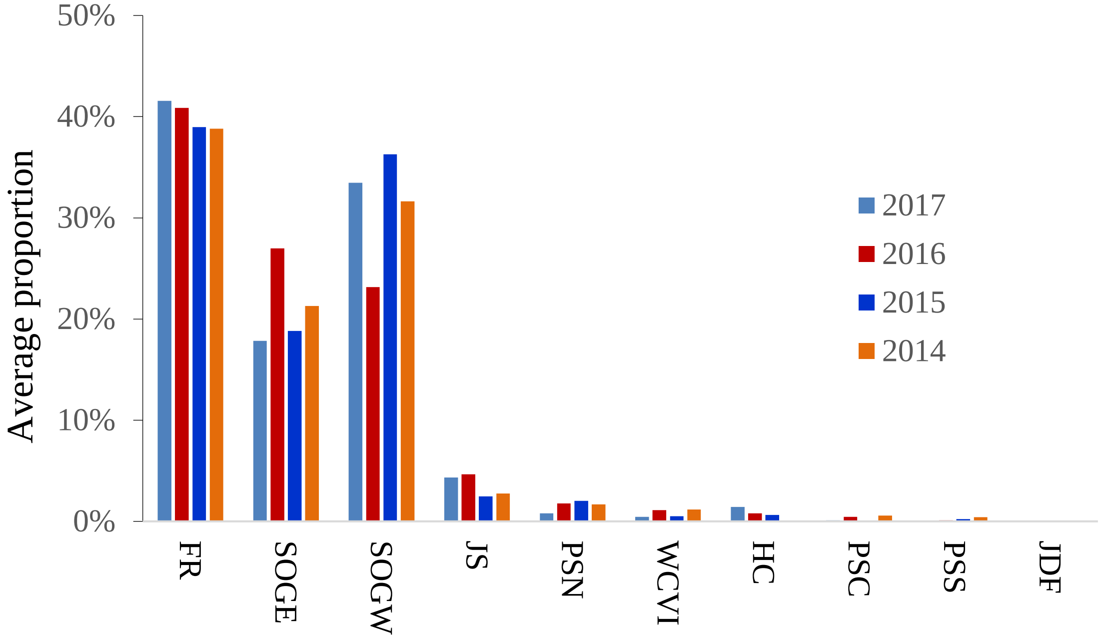

Figure 1. Average proportion of Genetic Units (GUs) in Canadian Management Area 12 from 2014 to 2017.

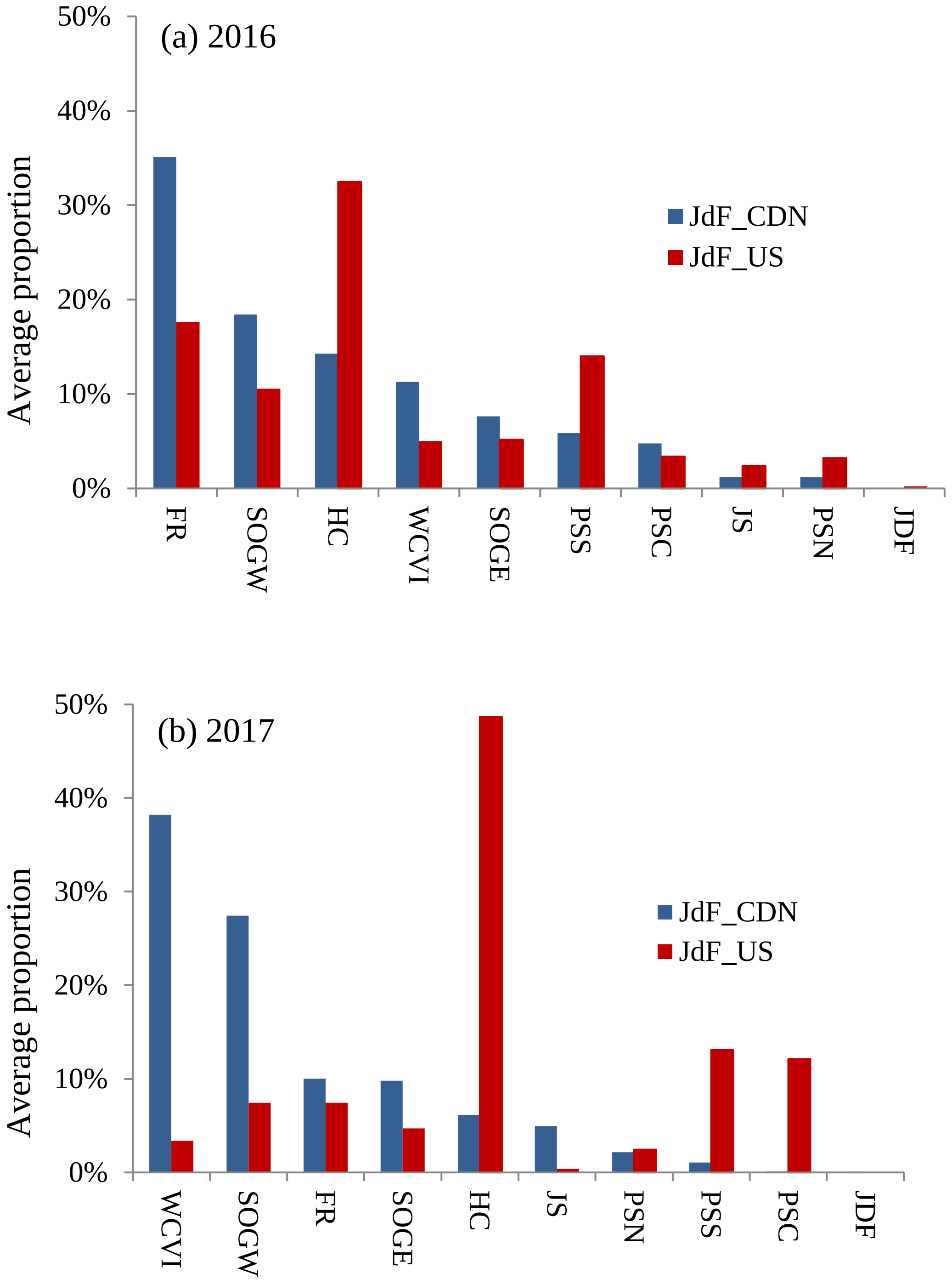

Figure 2. Average proportion of Genetic Units (GUs) in Canadian waters and U.S. waters of the Strait of Juan de Fuca (JdF) in (a) 2016 and (b) 2017.

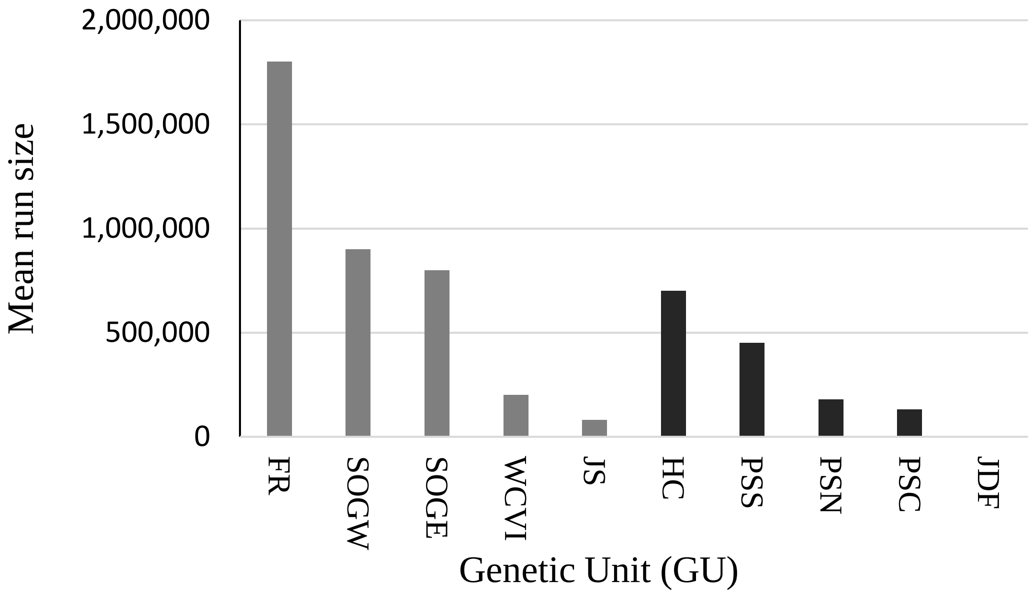

Figure 3. Priors of mean run size estimates for Canadian (grey) and U.S. (black) fall Genetic Units (GUs). The mean run size of the JdF GU is estimated as 2,000, too low to be shown accurately in the figure.

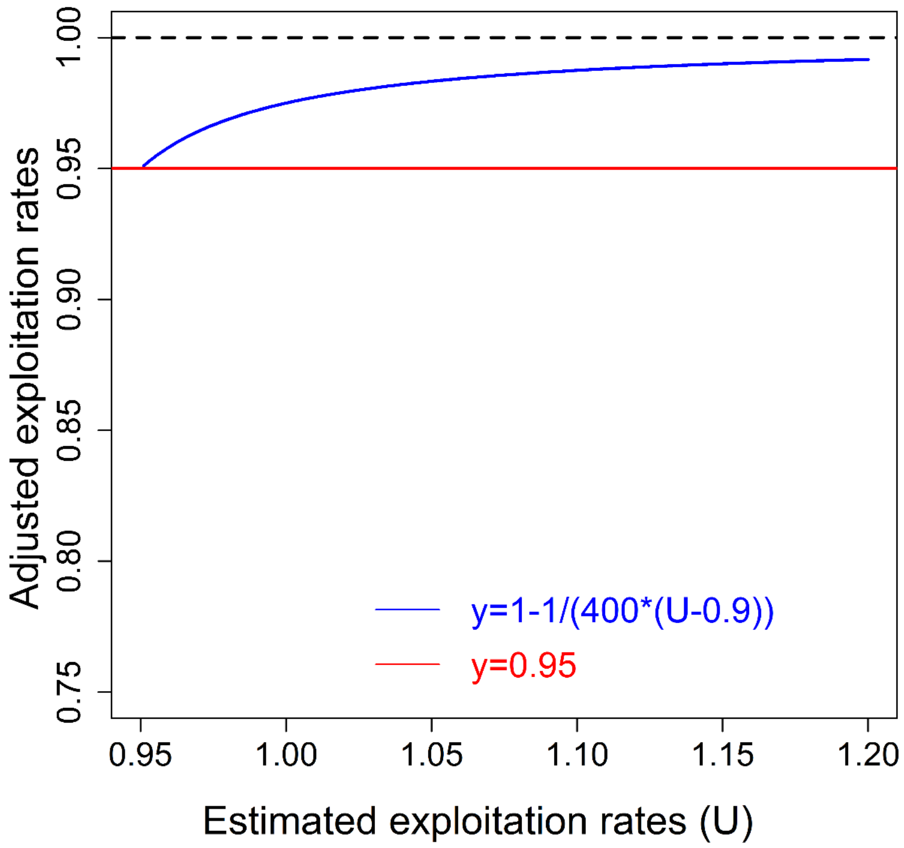

Figure 4. Adjusted exploitation rates when the estimated exploitation rates are greater than 0.95. The blue line shows the corrected function and the red line shows the original function with equations in the legend.