## UPdating FRASER RIVER SOCKEYE SALMON RUN SIZE EStimation USING BOOSTED Regression Tree Method

Yi Xu1, Mike Hawkshaw1, Caihong Fu2, Roy Hourston3, David Patterson4, Peter Chandler3

1Fisheries and Oceans Canada, Fraser and Interior Area, Delta, BC, [yi.xu2@dfo-mpo.gc.ca](mailto:yi.xu2@dfo-mpo.gc.ca), [mike.hawkshaw@dfo-mpo.gc.ca](mailto:mike.hawkshaw@dfo-mpo.gc.ca)

2Fisheries and Oceans Canada, Pacific Biological Station, Nanaimo, BC, [caihong.fu@dfo-mpo.gc.ca](mailto:caihong.fu@dfo-mpo.gc.ca)

3Fisheries and Oceans Canada, Institute of Ocean Sciences, Sidney, BC, [roy.hourston@dfo-mpo.gc.ca](mailto:roy.hourston@dfo-mpo.gc.ca) [peter.chandler@dfo-mpo.gc.ca](mailto:peter.chandler@dfo-mpo.gc.ca)

4Fisheries and Oceans Canada, Science Branch, Pacific Region, School of Resource and Environmental Management, Simon Fraser University, Burnaby, BC, [david.patterson@dfo-mpo.gc.ca](mailto:david.patterson@dfo-mpo.gc.ca)

### Highlights

* A Boosted Regression Trees model (BRT) was developed to study the mathematical relationships between Fraser River Sockeye Salmon recruitment and multiple environmental variables.
* In general, the BRT models are able to reproduce major variations observed and can explain over 50% of the variability in recruitment of all selected Sockeye Salmon stocks.
* The BRT models identify effective females spawners or juvenile abundance as the top contributor for predicting the recruitments of 15 (out of 19) Sockeye Salmon stocks while the contributions of various environmental parameters are less preeminent (<30% of the total recruitment variance) and diverse in their relationships with the recruitments of the stocks.
* BRT forecasts of Sockeye Salmon recruitment are a viable alternative to current forecast models to inform stock assessment and harvest planning for the coming fishing season.

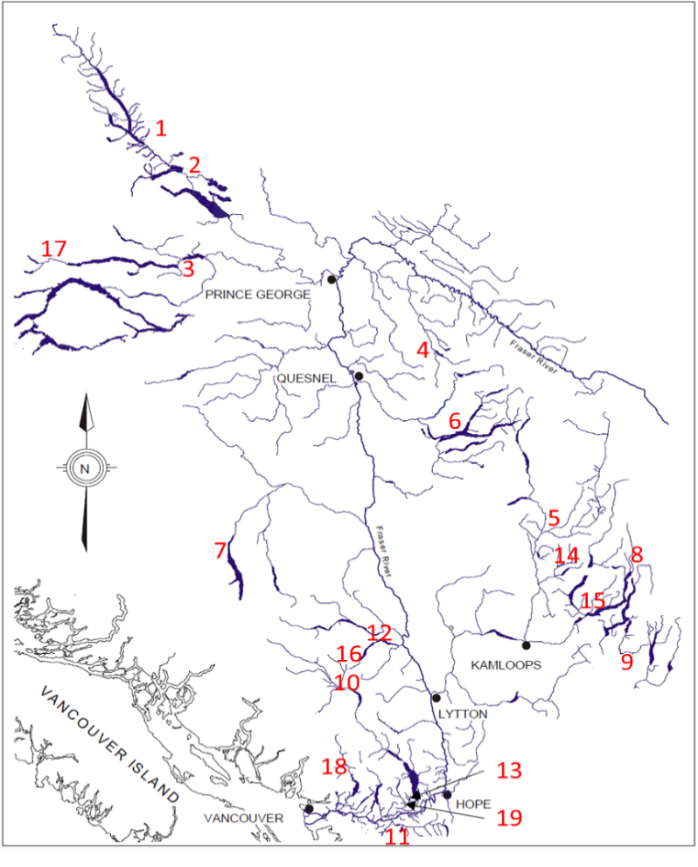
### Descriptions of data time series and the machine learning method

#### Fish Population Data

The Fraser Sockeye Salmon (*Oncorhynchus nerka*) recruitment time series (1948-2016) were provided by the Pacific Salmon Commission (PSC) and the time series of effective females spawners (EFS) and juvenile abundance (JUV) for the same period were provided by DFO for 19 major stocks (Figure 68-1). These data sets, detailed in Grant et al. (2011) were used to forecast Sockeye run size in 2020 (Hawkshaw et al. 2020).

#### Environmental Data

The 2019 forecast models incorporated time series of the Pacific Decadal Oscillation (PDO, Nov -Mar), sea surface temperatures (SST) from Pine Island (Apr-Jul), Entrance Island (Apr-Jun), and Fraser River discharge (Apr-Jun) at Hope as environmental covariates. In this new forecast, we added additional oceanographic variables and climate indices as candidate covariates (Table 68-1). The oceanographic variables include: the averaged SST of the Gulf of Alaska from the Centennial *in-situ* Observation-Based Estimates model (COBE; Ishii et al. 2005), and regional upwelling and downwelling favourable wind stress (Kistler et al. 2001; Hourston and Thomson, Section 8). The climate indices considered are: the seasonal and annual North Pacific Gyre Oscillation (NPGO; Di Lorenzo et al. 2008), the Northern Oscillation Index (NOI; Schwing et al. 2002) and the North Pacific Current Bifurcation Index (BI; Cummins and Freeland 2007). The time series of these all the variables are from 1950-2018 except for BI (from 1967-2018).



*Figure 68-1. Locations of 19 major Fraser Sockeye salmon stocks where spawning data were collected.*

1.Early Stuart

2.Late Stuart

3.Stellako

4.Bowron

5.Raft

6.Quesnel

7.Chilko

8.Seymour

9.Late Shuswap

10.Birkenhead

11.Cultus

12.Portage

13.Weaver Creek

14.Fennel Creek

15.Scotch Creek

16.Gates

17.Nadina

18.Upper Pitt River

19.Harrison

*Table 68-1. Leading environmental factors identified by Boosted regression trees model.*

|  |  |  |
| --- | --- | --- |
| Climate indices | Spring PDO (Pacific Decadal Oscillation Mar, Apr, May)  Summer and Autumn NPGO (Jun, Jul, Aug)  Winter NOI (Northern Oscillation Index)  Bifurcation Index | pdo.spr  npgo.aut, npgo.sum  noi.win  BI |
| Regional sea surface temperature and salinity | Averaged Entrance Island and Pine Island SST  Monthly Entrance Island SST(Apr, Jun)  Monthly Pine Island April SST (Apr, May, Jun, Jul)  Annual Gulf of Alaska SST  Summer Gulf of Alaska SST(Jun, Jul, Aug)  Autumn Gulf of Alaska SST(Sep, Oct, Nov)  Averaged Race Rocks and Amphitrite Point SSS (Jul, Aug, Sep) | eisst, pisst  apesst, jnesst  appsst, mapsst, jnpsst  jlpsst  ocean.sst.annual  ocean.sst.sum  ocean.sst.aut  stn2js |
| Wind stress | Annual,  Spring upwelling-favoured at Central Coast  Summer downwelling-favoured at Central Coast  Annual, autumn and winter downwelling-favoured at Prince Rupert District  Annual and winter downwelling-favoured at Prince Rupert District | wind.cc.up.annual  wind.cc.up.spr  wind.cc.dn.sum  wind.prd.dn.annual  wind.prd.dn.aut wind.prd.dn.win  wind.prd.up.annual  wind.prd.up.win |

#### Boosted regression trees model

A boosted regression trees (BRT, Elith et al. 2008) model was developed to study the mathematical relationships between Sockeye recruitment and multiple environmental co-variates since 2019 (Xu et al., 2019; 2020). This model is based on a machine learning method and has three advantages: 1) it can fit complex nonlinear relationships easily with multiple predictors; 2) it is not sensitive to outliers and data transformation; and 3) it is able to handle missing data. The BRT model was implemented using packages of “gbm” (generalized boosted regression models, v2.1.5) and “dismo” (species distribution modeling, v1.1-4) in R (R Development Core Team 2021). No major model changes have been made since last year.

### Status and trends

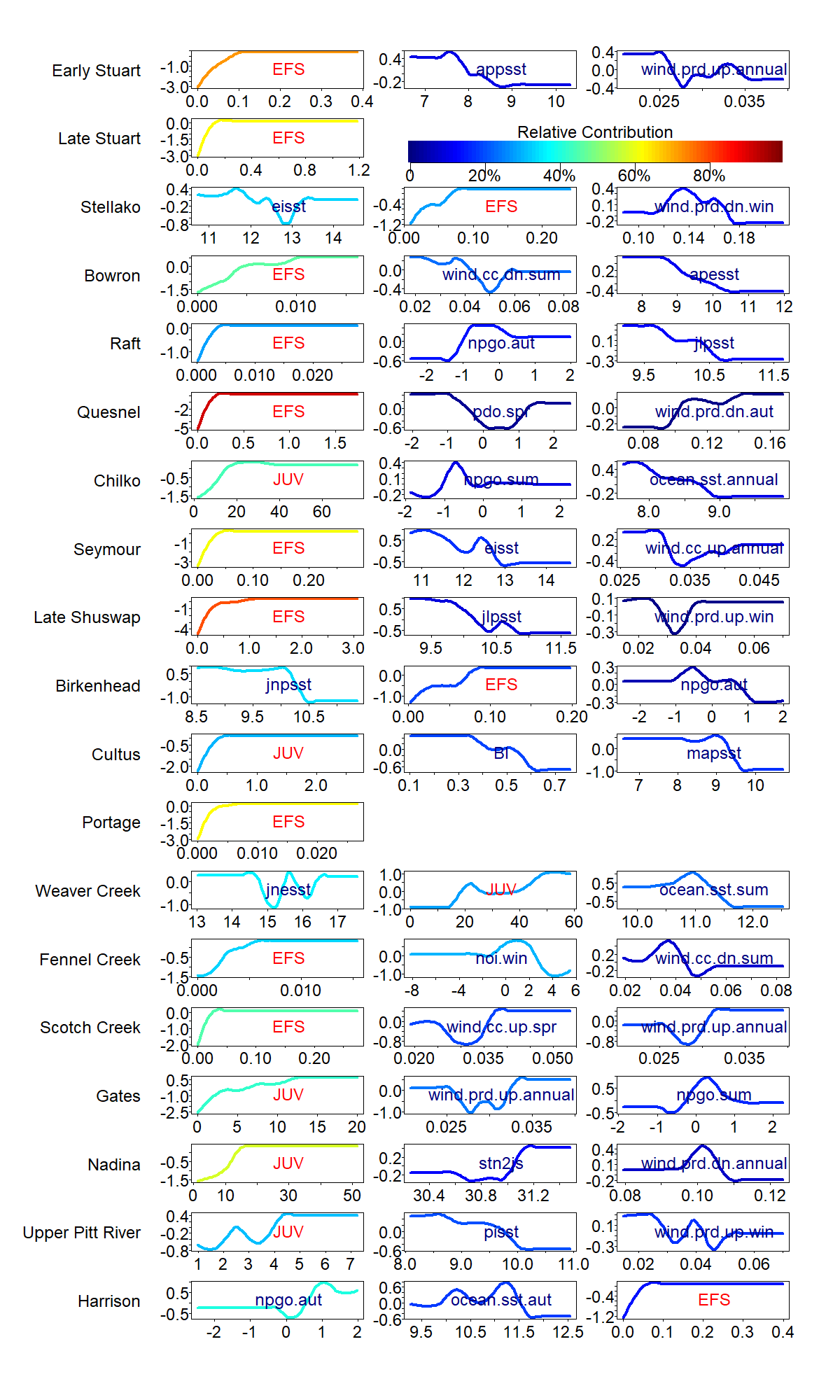
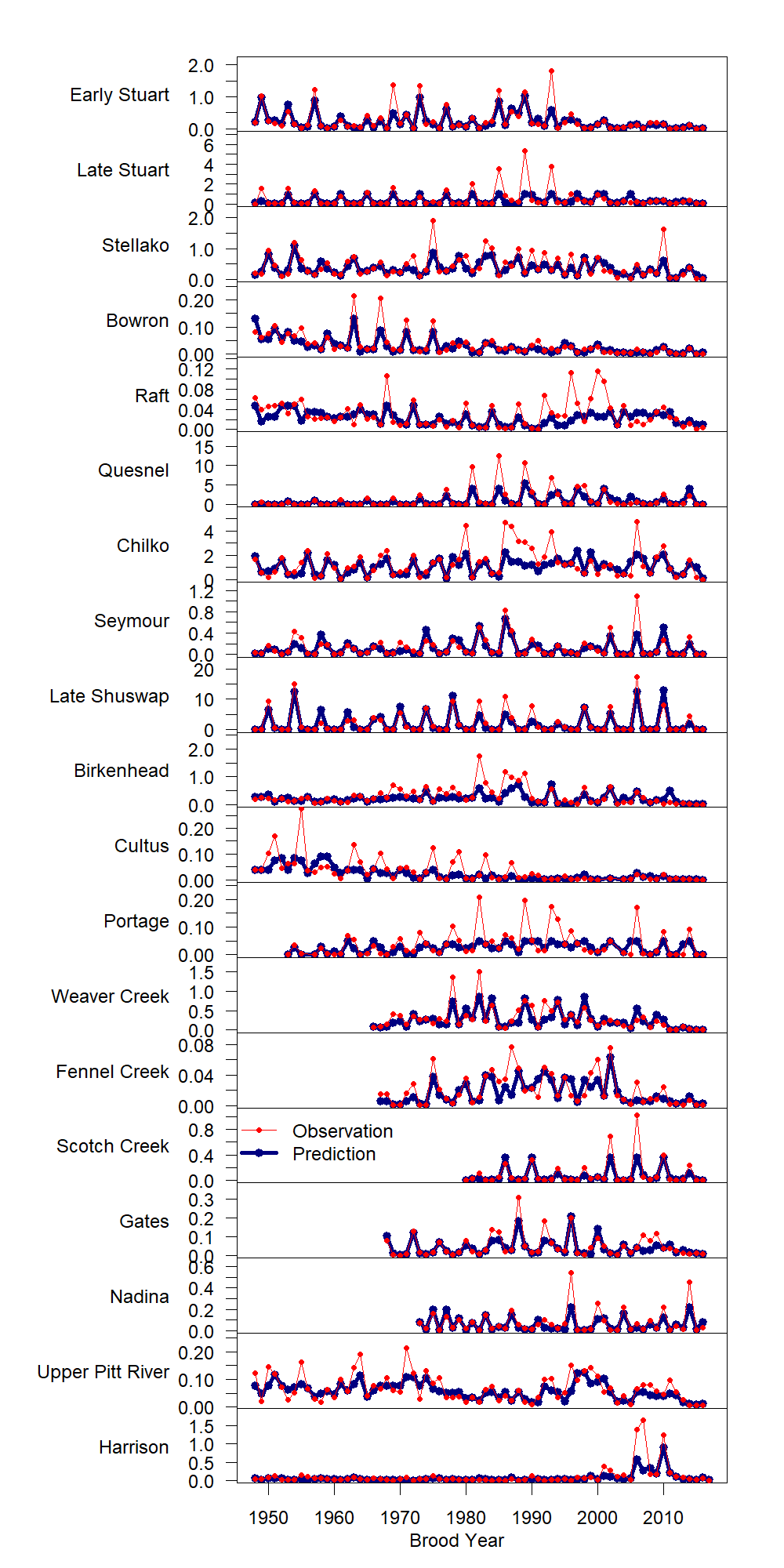
Since the late 1990s, most Fraser River Sockeye Salmon stocks have experienced low recruitment. In general, the BRT model fit was able to explain a large proportion of the variability in the recruitment time series (Figure 68-2) with the Quesnel stock achieving the highest level (93.7%). For the majority of stocks the BRT model was able to predict the general recruitment trends, although missing some extremes in the observed values, which in turn was reflected in the relatively small standard deviations in the predictions.

### Factors influencing trends

For most Fraser River Sockeye Salmon stocks, the BRT models identified EFS or JUV as the top contributor that had the highest relative influence (%) for predicting Sockeye Salmon recruitment (Figure 68-3). For all stocks (except for the Weaver Creek – Ricker?), predicted recruitment showed a Beverton-Holt-like relationship with EFS or JUV, and the relative influences varied from 14-91% among different stocks. While the relationship between recruitment and the top biological factor (EFS or JUV) was shown as Beverton-Holt-like, the relationship between recruitment and the dominant environmental factor was diverse in shapes. Environmental factors explained less than 30% of the total recruitment variance. For stocks where a dominant environmental factor was identified as the top contributor (i.e., the Stellako, Birkenhead, Weaver Creek and Harrison stocks), the dominant environmental factor showed a smaller contribution (with lower relative influence) compared to a top biological factor in other stocks.

### Run size forecasts and implication of the BRT modelling method

The BRT model produced forecasts of Sockeye Salmon run size for 19 major stocks (Figure 68-4) totaling around 2 million in the year 2021. The Late Stuart (907K), Quesnel (281K) and Chilko (720K) stocks dominate the run size and represent 90% of the forecasted total of all stocks combined. However, for both Late Stuart and Chilko stocks, the explained variance of BRT model is relative low (less than 70%), which indicates higher uncertainties of the forecast. In contrast, all other stocks (Early Stuart, Stellako, Bowron, Raft, Seymour, Late Shuswap, Birkenhead, Cultus, Portage, Weaver Creek, Fennel Creek, Scotch Creek, Gates stocks, Upper Pitt River and Harrison) are predicted less than 75,000 recruits; this is the abundance threshold guideline for determining whether high precision spawning escapement methods (e.g. sonar, mark-recapture) should be planned for the upcoming year. These stock-specific results provide useful and timely information to both stock assessment and harvest managers for the upcoming 2021 summer/fall enumeration surveys.

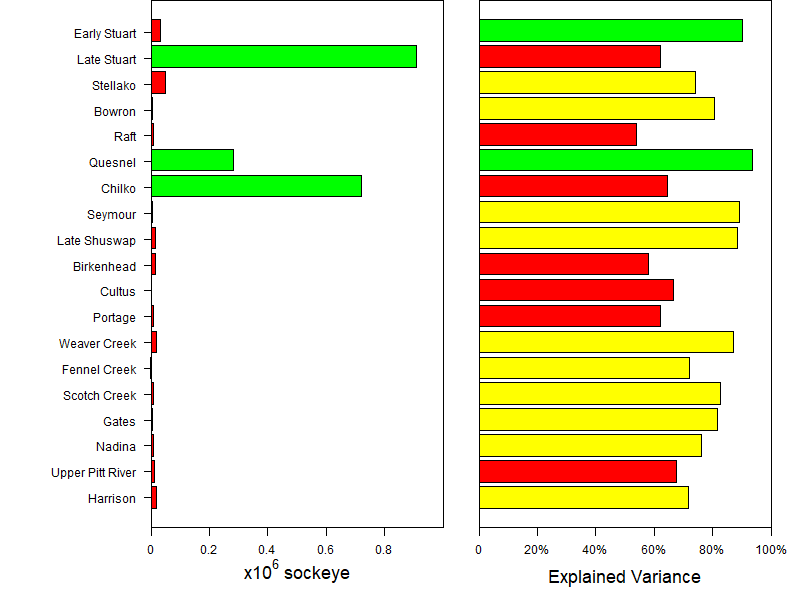


*Figure 68-3. Fitted functions of top three predictors and relative contributions from the Boosted Regression Trees models (See Table 68-1 and text for acronym definitions).*

*Figure 68-2. Observed and Boosted-Regression-Trees predicted recruitment (log scale) of 19 Fraser River Sockeye Salmon stocks.*

*Figure 68-3. Fitted functions of top three predictors and relative contributions from the Boosted Regression Trees models (See Table 68-1 and text for acronym definitions).*

*Figure 68-2. Observed and Boosted-Regression-Trees predicted recruitment (log scale) of 19 Fraser River Sockeye Salmon stocks.*



*Figure 68-4. Left: The Boosted Regression Trees model run size forecasts of 19 Fraser River Sockeye salmon stocks in 2020 (Red: <=75,000, Green: >75,000).Right: The explained variance by Boosted Regression Tree model (Red <=70%, Yellow 70-90%, Green >=90%).*

### References

Cummins, P.F., and Freeland, H.J. 2007. Variability of the North Pacific Current and its bifurcation. Progress in Oceanography 75: 253-265.

Di Lorenzo, E., Schneider, N., Cobb, K.M., Chhak, K, Franks, P.J.S., Miller, A.J., McWilliams, J.C., Bograd, S.J., Arango, H., Curchister, E., Powell, T.M., and Rivere, P. 2008. North Pacific Gyre Oscillation links ocean climate and ecosystem change. Geophysical Research Letters 35: L08607.

Elith, J., Leathwick, J.R., and Hastie, T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77: 802-813.

Grant, S.C.H., MacDonald, B.L., Cone, T.E., Holt, C.A., Cass, A., Porszt, E.J., Hume, J.M.B., and Pon, L.B. 2011. Evaluation of uncertainty in Fraser Sockeye (*Oncorhynchus nerka*) Wild Salmon Policy status using abundance and trends in abundance metrics. Can. Sci. Advis. Sec. Res. Doc. 2011/087: viii + 183 p.

Hawkshaw, M., Xu, Y. and Davis, B. 2020. Pre-season run size forecasts for Fraser River sockeye (Oncorhynchus nerka) salmon in 2020. Can. Tech. Rep. Fish. Aquat. Sci. 3392: vi. +52p.

Ishii, M., Shouji, A., Sugimoto, S., and Matsumoto, T. 2005. Objective Analyses of Sea-Surface Temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe collection. Int. J. Climatol. 25: 865-879.

Kistler, R., Kalnay, E., Collins, W., Saha, S., White, G., Woollen, J., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., van den Dool, H., Jenne, R., and Fiorino, M. 2001. The NCEP–NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. Bulletin of Amerian Meteorological Society 82: 247–268.

R Development Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL. <https://www.R-project.org>

Xu, Y., M. Hawkshaw, C. Fu, D. Patterson, R. Hourston, P. Chandler. Estimating Fraser River sockeye salmon run size using a machine learning method. In Boldt, J.L., Javorski, A., and Chandler, P.C. (Eds.). 2020. State of the physical, biological and selected fishery resources of Pacific Canadian marine ecosystems in 2019. Can. Tech. Rep. Fish. Aquat. Sci. 3377: x + 288 p.

Xu, Y., M. Hawkshaw, D. Patterson, R. Hourston, P. Chandler. How Fraser sockeye salmon recruitment was affected under climate changes? A model study. In Boldt, J.L., Leonard, J., and Chandler, P.C. (Eds.). 2019. State of the physical, biological and selected fishery resources of Pacific Canadian marine ecosystems in 2018. Can. Tech. Rep. Fish. Aquat. Sci. 3314: vii + 248 p.