Data from: Iacarella, JC and Weller, JD (2023) Predicting favourable streams for anadromous salmon spawning and natal rearing under climate change. *Canadian Journal of Fisheries and Aquatic Sciences*, see CJFAS for volume and page numbers.

Please see the paper for method explanations.

<u>**Geodatabases**</u>

1. **'FullExtent_accessible_bySpecies.gdb'**

This geodatabase contains the freshwater network used for fitting Environmental Niche Models to predict salmon spawning habitat in BC and western US salmon bearing watersheds. 'Reaches' are the linear features, and 'midpoints' are the central points of each reach. See shapefile metadata for attribute details.

The 'midpoints' file was used to make environmental rasters. These rasters were then used in modeling, and later for remapping the model outputs.

2. **'Salmon_spawning_occurrence.gdb'**

Stream reach-level observations of spawning locations for each salmon species, appended to the midpoints (i.e. central points) of the stream reaches.

3. **'BC_network_variables_climateproj.gdb'**

This geodatabase contains the freshwater network used for model projections in BC salmon bearing watersheds. 'Reaches' are the linear features, and 'midpoints' are the central points of each reach. See shapefile metadata for attribute details.

The 'midpoints' file was used to make environmental rasters. These rasters were then used in modeling, and later for remapping the model outputs.

4. **'Mapped_model_predictions.gdb'**
   - Generalized Boosted Model favourability predictions for each species' extent, accessible reaches only, baseline conditions: 'Species common name_GBMFav_streams_acc'
   - Generalized Boosted Model change in favourability predictions for BC streams reaches within salmon bearing watersheds, comparing favourability from 1981-2000 to 2081-2100 under three climate change scenarios:
     - Ensemble GCM, RCP 2.6, end of century (2081-2100): 'SPECIES COMMON NAME_BC_9_26_5'
     - Ensemble GCM, RCP 4.5, end of century (2081-2100): 'SPECIES COMMON NAME_BC_9_45_5'
     - Ensemble GCM, RCP 8.5, end of century (2081-2100): 'SPECIES COMMON NAME_BC_9_85_5'

Spreadsheets

1. 'SPECIES COMMON NAME_FullExtent_accessible_data_preds_PresentConditions_072023.csv'

These CSVs contain presence-absence identification and environmental variables used to fit models for accessible reaches across the full study extent, as well as model outputs.

Model predictions were mapped by first converting the prediction value of interest to a raster (by snapping to a related environmental raster), extracting the values to the related 'midpoints' network file, and then joining to the 'reaches' file.

Column names are:

'Presence' – presence/absence identification

'x', 'y', 'cells' used to grid the presence/absence data with the environmental rasters

'ELEV' – reach elevation (m)

'GRAD' – reach gradient (m/m)

'TEMP' – August mean stream temperature (°C) at the reach

'PPT' – Annual precipitation (mm) at the reach

"AREA' – Catchment area of stream reach (hectares)

'OCEAN' – distance (m) to the nearest coastline

'LAKE' (Sockeye only) – distance (m) to the closest accessible lake using the full stream network

'GAM_P' – probability predictions from Generalized Additive Models

'MX_P' – probability predictions from Maxent

'RF_P' – probability predictions from Random Forest models

'GBM_P' – probability predictions from Generalized Boosted Models

'BART_P' – probability predictions from Bayesian Additive Regression Trees

'BART_Fav' – favourability predictions from Bayesian Additive Regression Trees

'GBM_Fav' – favourability predictions from Generalized Boosted Models

2. 'SPECIES COMMON NAME_Proj_all_FavChange_GBM_072023.csv'

These CSVs contain probability and favourability values from Generalized Boosted Models projected under different climate change scenarios and time steps.

Model predictions were mapped by first converting the prediction value of interest to a raster (by snapping to a related environmental raster), extracting the values to the related 'midpoints' network file, and then joining to the 'reaches' file.

Column names are:

'Presence' – presence/absence identification

'x', 'y', 'cells' used to grid the presence/absence data with the environmental rasters

'SDM_ID_all' – stream reach ID corresponding with GDB files

Prob.f…': Probability predictions for each climate scenario and time step using Generalized Boosted Models

Fav.f…': Favourability predictions for each climate scenario and time step using Generalized Boosted Models

Naming structure for probability and favourability columns: '3-character attribute code'_'1-digit GCM code'_'2-digit_RCP code'_'1_digit time period code';

GCM codes:

0 = historic data, 1 = CanESM2, 2 = CSIRO, 3 = GFDL, 4 = HadGEM, 5 = MIROC, 6 = MPI, 9 = ensemble mean of models 1-6

RCP codes:

00 = historic data, 26 = RCP 2.6, 45 = RCP 4.5, 85 = RCP 8.5

Time period codes:

0 = 1981-2000, 1 = 2001-2020, 2 = 2021-2040, 3 = 2041-2060, 4 = 2061-2080, 5 = 2081-2100