# Forecasting Model Workflow Instructions

**Subject:** Run Forecasting for Hood River Spring Chinook and Steelhead

**Date:**

**Prepared for:** The Confederated Tribes of the Warm Springs

**Prepared by:** Mount Hood Environmental

# Overview

This document provides a step-by-step guide to update and run forecasting models for Hood River spring Chinook and steelhead. This is intended to be a user-friendly workflow that requires a minimum amount of data updates and technical ability with statistical software. After updating data inputs and running the R script, a report will be generated that provides a table of forecasts along with model fit statistics.

# Required and Recommended Items

1. A copy of the *RunForecast\_HoodRiver* folder, which contains the necessary scripts and data to run the models. This folder can be placed in any easily accessible directory on the computer.
2. An up-to-date installation of the statistical program R
   1. Navigate to <https://cran.r-project.org/>, follow instructions to download and install for the appropriate operating system.
3. An installation of RStudio is recommended, but not required. RStudio creates a more user-friendly interface for R.
   1. RStudio can be downloaded and installed from <https://posit.co/download/rstudio-desktop/>
4. An internet connection while running the report script.
5. Updated input data for predictors used in the model. This will be covered below.

# Updating Data Inputs

Each year, predictor data will need to be updated to produce forecasts. This includes prior abundance data such as returns, smolt outmigration, and counts from Bonneville Dam and the Dalles. All data inputs are housed within the *data input* folder.

Adult Abundance Data

Abundance data is housed within the *AbundanceData* spreadsheet located at *RunForecast\_HoodRiver/data input/Abundance/AbundanceData*. Within this spreadsheet, there are tabs for adult and smolt abundance data.

In the adult data spreadsheet, the following columns will need to be updated:

* Year: Populate a new row in this column with the previous year (e.g., when forecasting for 2023, enter 2022).
* *SPCH BON*
  + This is the abundance of spring Chinook passing Bonneville Dam in a given return year.
* *SPCH DALLES*
  + This is the abundance of spring Chinook passing the Dalles in a given return year.
* *SSTHD BON*
  + This is the abundance of summer steelhead passing Bonneville Dam in a given return year.
* *SSTHD DALLES*
  + This is the abundance of summer steelhead passing the Dalles in a given return year.

These data can be obtained from Columbia River DART (<https://www.cbr.washington.edu/dart>). Starting with the spring Chinook data, navigate to: <https://www.cbr.washington.edu/dart/query/adult_annual_sum> and ensure that the following criteria are selected (Figure 1):

1. **Select Output Format:** *HTML Table with CSV File*
2. **Select Project:** *BON-Bonneville*
3. Under **Set Date Range:** Select *Restrict by Chinook Run Dates* to *Spring*
4. Submit query

This will bring you to a page with a table of query results. Select the count value for the most recent year in the *Chinook* column, then paste into the *SPCH BON* column of the *AbundanceData* spreadsheet within the *Adult* tab for the appropriate year. Repeat this process with the **Select Project** field as *TDA-The Dalles*, and enter this value into the *SPCH DALLES* column.

To update the *SSTHD BON* and *SSTHD DALLES* columns, we will perform a similar process.

1. **Select Output Format:** *HTML Table with CSV File*
2. **Select Project:** *BON-Bonneville*
3. Under **Set Date Range:** Select **Start** = 5/1, **End =** 9/30
4. Submit query

This provides an abundance value for the *SSTHD BON* column. Once again, repeat for the Dalles to get a value and populate the *SSTHD DALLES* column.

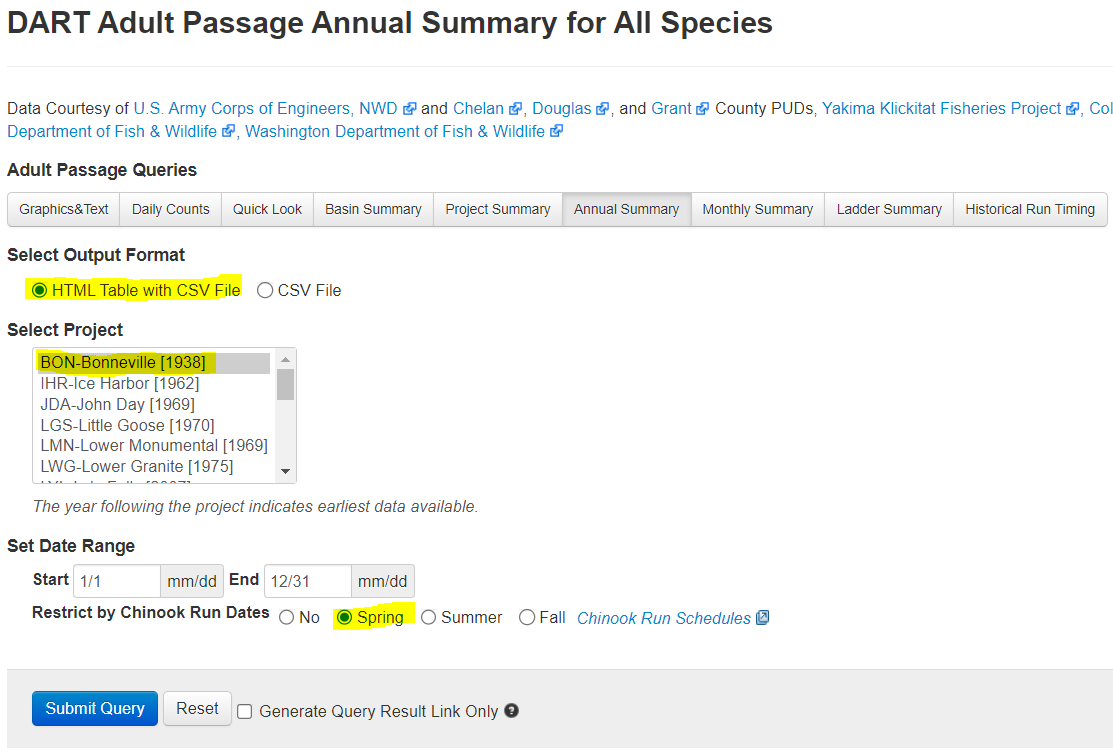


Figure . Example query for generating spring Chinook counts at Bonneville Dam for the “SPCH BON” column.

Smolt Abundance Data

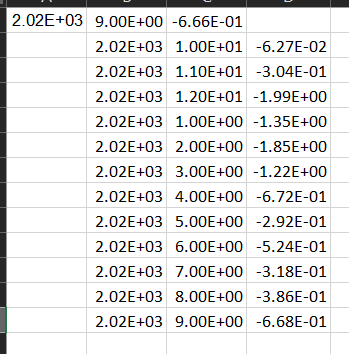
Within the *AbundanceData* spreadsheet, navigate to the *Smolt* tab. Here, we will need to update the *Year*, *HOR WSTHD*, *HOR SPCH*, and *NOR STHD* columns with the most recently available data.

* *Year*: Populate a new row in this column with the previous year (e.g., when forecasting for 2023, enter 2022).
* *HOR WSTHD*: Enter a zero value, as the program has ceased operation
* *HOR SPCH*: Enter the number of hatchery smolts released, obtained from FLR slips
* *NOR STHD*: Enter the estimated abundance of NOR steelhead smolt outmigrants for the given year.

This completes all the updates required to the *AbundanceData* file. Please save and close the spreadsheet before continuing.

Environmental Data

The models utilize ocean condition and Hood River flow data. Hood River flow data is automatically obtained during script operation, so the only update required is for NPGO (Northern Pacific Gyre Oscillation).

1. Open the *NPGO* spreadsheet, located at *RunForecast\_HoodRiver/data input/Environmental/NPGO*
   1. Scroll to the bottom of this spreadsheet to see when the data ends. As of writing, NPGO is updated to September 2022.
2. Navigate to <http://o3d.org/npgo/>
   1. Click *Download NPGO Index* on the webpage’s header ribbon. This will bring you to a plain text page with a ton of NPGO data. All of the values are in scientific notation, but we can handle that.
3. Scroll to the bottom of the page to get values for the most recent year.
   1. Highlight and copy all the new data available. The leftmost column is year, and the middle column is month. Since we’re in scientific notation, September 2022 looks like “2.0220000e+03 9.0000000e+00”.
4. Paste the new data into the leftmost column on the bottom of the data in the *NPGO* spreadsheet.
5. Ok, now a little formatting.
   1. Highlight the left column where all the new data is. It’s all going to be stuck within one column, so let’s fix that.
   2. Next, click on the *Data* ribbon in Excel, and then click on the *Text to Column* button, located within the *Data Tools* box.
   3. This will give a pop-up. For the first page, we choose *Delimited* and click next. On the next page, under **Delimiters,** check *space* and uncheck all other boxes. Then click finish.
   4. This fixes most of it, but the data is probably not in three columns like we want it to be, thanks Excel!
      1. If your data looks like this: 
         1. Then go ahead and highlight all that empty space after the first row in the left column, right click, click on *Delete* and inside the pop-up choose *Shift cells left* and click ok.
   5. Now if we want to deal with that nasty scientific notation, highlight all the new data, navigate to the *Home* ribbon in Excel, and change the type from *Scientific* to *General* or *Numeric* in the drop-down menu.
6. Save and close this spreadsheet.

# Running the Report

Now that the data is updated, we can go ahead and run the models and report.

1. Within the main *RunForecast\_HoodRiver* folder, open the *RunForecast\_HoodRiver* R Project file. This will open in R, or RStudio if installed.
2. Once the R project opens, navigate to *file* → *open file* (or ctrl + o), and open the *Generate\_report* R script located within the main *RunForecast\_HoodRiver* folder.
3. Update the *prediction\_year* as needed.
   1. On line 6, update the *prediction\_year = YYYY* statement to the appropriate year.
4. Highlight the script and click *Run* (or ctrl + enter).

This will generate a forecast report for the desired year and house the report and figures within directories for that year. The report will be output as an .html file located in a folder for that year within the *report output* folder. It will also produce figures for model observed and predicted values within the “figures” folder. **Note:** This process will take a few minutes to fit new models before generating the report. Additionally, if this is the first time running the report on a computer, it will take a few minutes to download and install the necessary R packages.

This concludes the basic workflow for updating data inputs and running forecast predictions. However, there may (and should be!) a desire to re-fit these models with new returns data as it becomes available, even if it is not possible for all runs. Details for this process are below.

# Fitting Models with Updated Returns Data

When we get some new returns data, let’s incorporate it and fit models with the updated data.

1. Open up the *AbundanceData* spreadsheet
2. On the *Adult* tab, the following columns relate to Hood River returns:
   1. *NOR WSTHD*: This is the abundance of returning NOR winter steelhead, and can be updated with additional data for future run years.
   2. *WSTHD total escapement*: Data in this column was used to estimate smolt abundance for 2021 and 2022, and does not need to be updated as these estimates have already been calculated.
   3. *HOR WSTHD*: Abundance of HOR winter steelhead, can be updated with new returns, however the utility of updating this column is limited due to the end of the hatchery program.
   4. *SSTHD*: Abundance of summer steelhead, can be updated for future run years.
   5. *NOR SPCH* and *NOR SPCH JACK*: NOR spring Chinook adult and jack abundances, updating this data with return abundance when possible is advised, as the current datasets are quite limited.
   6. *HOR SPCH* and *HOR SPCH JACK*: HOR spring Chinook adult and jack abundances. Updating these data is recommended when available.
3. Update number of the above columns with new returns data, making sure to populate the *Year* column as well. Please ensure that years are consecutive without gaps.
4. Save and close out the spreadsheet, then run the report. **Note:** If the report has already been run for the year, you will need to delete the model objects that were created in order to fit models with the new data.
   1. To do so, navigate to *R/model objects/* and select the folder with the prediction year.
   2. Inside the appropriate year folder, delete all the files. They will be regenerated when the report is run.