

Appendix S1. Instructions for retrieving and archiving the environmental covariates.

Contents

Background	1
User inputs	2
Retrieve covariates	2
River discharge	2
Winter maximum	4
Summer minimum	5
This is now Spring/Summer Maximum from May through July	5
North Pacific Gyre Oscillation	5
Hatchery releases	7
Archive covariates	7

This is version 0.23.02.17.

Background

This appendix describes how to retrieve the environmental covariates used in our analyses. After reading in the raw data, summarizing them (if necessary), and trimming them to the appropriate time frame, the table of covariates is written to a `.csv` file.

All of the analyses require the R software (v3.4.3 or later) for data retrieval and processing. We also need the **readr** and **here** packages, which are not included with the base installation of **R**.

```
## load libraries
library("readr")

library("here")

## set data dir
datadir <- here("data")
```

User inputs

We begin by supplying values for the following parameters, which we use for trimming and lagging the covariates to the appropriate years.

```
## first & last years of fish data
yr_first <- 1978
yr_last <- 2018

## min & max adult ages (years)
age_min <- 3
age_max <- 8

## time lags (years) for covariates
flow_lag <- 0 #change lag to zero so that flow data aligns with fry stage
marine_lag <- 2
hrel_lag <- 2

## number of years for run forecasts
n_fore <- 0
```

Retrieve covariates

Our analysis investigates 4 covariates as possible drivers of the population's intrinsic growth rate:

1. Maximum river discharge in winter;
2. Minimum river discharge in summer;
3. North Pacific Gyre Oscillation;
4. Releases of hatchery-born juveniles.

River discharge

We begin by getting the daily flow data from the US Geological Service National Water Information System. We will use the direct link to the gage data from the Skagit River near Mount Vernon, WA (#12178000), beginning with the first year of fish data. #this gauge is for Newhalem not Mount Vernon??

```
## flow gage ID
flow_site <- 12181000
## get URL for flow data from USGS
flow_url <- paste0("https://waterdata.usgs.gov/nwis/dv",
                  "?cb_00060=on", #This is param code 00060 for daily mean stream flow
                  "&format=rdb",
                  "&site_no=",flow_site,
                  "&begin_date=",yr_first,"-01-01",
                  "&end_date=",yr_last,"-12-31")
```

Next we retrieve the raw data file and print its metadata.

```
## raw flow data from USGS
flow_raw <- read_lines(flow_url)
## lines with metadata
hdr_flow <- which(lapply(flow_raw, grep, pattern = "\\#")==1, arr.ind = TRUE)
## print flow metadata
print(flow_raw[hdr_flow], quote = FALSE)

## [1] # ----- WARNING -----
## [2] # Some of the data that you have obtained from this U.S. Geological Survey database
## [3] # may not have received Director's approval. Any such data values are qualified
## [4] # as provisional and are subject to revision. Provisional data are released on the
## [5] # condition that neither the USGS nor the United States Government may be held liable
## [6] # for any damages resulting from its use.
## [7] #
## [8] # Additional info: https://help.waterdata.usgs.gov/policies/provisional-data-statement
## [9] #
## [10] # File-format description: https://help.waterdata.usgs.gov/faq/about-tab-delimited-out
## [11] # Automated-retrieval info: https://help.waterdata.usgs.gov/faq/automated-retrievals
## [12] #
## [13] # Contact: gs-w_waterdata_support@usgs.gov
## [14] # retrieved: 2023-02-17 10:53:06 EST (vaww02)
## [15] #
## [16] # Data for the following 1 site(s) are contained in this file
## [17] # USGS 12181000 SKAGIT RIVER AT MARBLEMOUNT, WA
## [18] # -----
## [19] #
## [20] # Data provided for site 12181000
## [21] # TS parameter statistic Description
## [22] # 149352 00060 00003 Discharge, cubic feet per second (Mean)
## [23] #
## [24] # Data-value qualification codes included in this output:
## [25] # A Approved for publication -- Processing and review completed.
## [26] # e Value has been estimated.
## [27] #
```

Lastly, we extract the actual flow data for the years of interest and inspect the file contents.

```
## flow data for years of interest
dat_flow <- read_tsv(flow_url,
                     col_names = FALSE,
                     col_types = "ciDdc",
                     skip = max(hdr_flow)+2)
colnames(dat_flow) <- unlist(strsplit(tolower(flow_raw[max(hdr_flow)+1]),
                                     split = "\\s+"))
head(dat_flow)
```

```
## # A tibble: 6 x 5
##   agency_cd site_no datetime '149352_00060_00003' '149352_00060_00003_cd'
##   <chr>      <int> <date>          <dbl> <chr>
## 1 USGS      12181000 1978-01-01          7430 A
## 2 USGS      12181000 1978-01-02          7170 A
## 3 USGS      12181000 1978-01-03          8070 A
## 4 USGS      12181000 1978-01-04          7390 A
## 5 USGS      12181000 1978-01-05          7210 A
## 6 USGS      12181000 1978-01-06          8310 A
```

We only need the 3rd and 4th columns, which contain the date (`datetime`) and daily flow measurements (149352_00060_00003). We will rename them to `date` and `flow`, respectively, and convert the flow units from “cubic feet per second” to “cubic meters per second”.

```
## keep only relevant columns
dat_flow <- dat_flow[c("datetime", grep("[0-9]$", colnames(dat_flow), value = TRUE))]
## nicer column names
colnames(dat_flow) <- c("date", "flow")
## convert cubic feet to cubic meters
dat_flow$flow <- dat_flow$flow / 35.3147
## flow by year & month
dat_flow$year <- as.integer(format(dat_flow$date, "%Y"))
dat_flow$month <- as.integer(format(dat_flow$date, "%m"))
dat_flow <- dat_flow[, c("year", "month", "flow")]
```

Winter maximum

We are interested in the maximum of the daily peak flows from October through May during the first year that juveniles are rearing in streams. This means we need to combine flow values from the fall of year t with those in the winter and spring of year $t + 1$. We also need to shift the flow data forward by 0 year so they align with the juvenile life stage. Therefore, the flow time series will begin in 1978 and end in 2015.

```
## autumn flows in year t
flow_aut <- subset(dat_flow, (month >= 10 & month <= 12)
                  & year >= yr_frst & year <= yr_last-age_min+n_fore)
## spring flows in year t+1
flow_spr <- subset(dat_flow,
                  (month >= 1 & month <= 5)
                  & year >= yr_frst+flow_lag
                  & year <= yr_last-age_min+n_fore+flow_lag)
## change spr year index to match aut
flow_spr[, "year"] <- flow_spr[, "year"] - flow_lag
## combined flows indexed to brood year & calculate max flow
dat_flow_wtr <- aggregate(flow ~ year, data = rbind(flow_aut, flow_spr), max)
dat_flow_wtr[, "flow"] <- round(dat_flow_wtr[, "flow"], 1)
## change year index to brood year
```

```

dat_flow_wtr[, "year"] <- dat_flow_wtr[, "year"]
## for plotting purpose later
colnames(dat_flow_wtr)[2] <- "flow_wtr"

```

Summer minimum

Retrieving the minimum flow juveniles would experience during their first summer (June through September) is straightforward.

This is now Spring/Summer Maximum from May through July

There is now no time lag to align with spawn and emergence year

```

## summer flows in year t
flow_sum <- subset(dat_flow, (month>=5 & month<=7)
                  & year >= yr_first+flow_lag
                  & year <= yr_last-age_min+n_fore+flow_lag)
## change year index to brood year
flow_sum[, "year"] <- flow_sum[, "year"] - flow_lag
## combined flows indexed to brood year & calculate min flow
dat_flow_sum <- aggregate(flow ~ year, data = flow_sum, min)
dat_flow_sum <- round(dat_flow_sum, 2)
## for plotting purpose later
colnames(dat_flow_sum)[2] <- "flow_sum"

```

North Pacific Gyre Oscillation

We used the monthly NPGO data provided by Emanuele Di Lorenzo of the Georgia Institute of Technology, which are available [here](http://www.o3d.org/npgo/npgo.php). We begin by downloading the raw NPGO data and viewing the metadata.

```

## URL for NPGO data
url_NPGO <- "http://www.o3d.org/npgo/npgo.php"
## raw NPGO data
NPGO_raw <- read_lines(url_NPGO)
## line with data headers
hdr_NPGO <- which(lapply(NPGO_raw, grep, pattern="YEAR")==1, arr.ind = TRUE)
## print PDO metadata
print(NPGO_raw[seq(hdr_NPGO)], quote = FALSE)

## [1]
## [2] <html>
## [3] <body>
## [4]
## [5] <pre># Last update 03-Oct-2022 by E. Di Lorenzo

```

```

## [6] # NPGO index monthly averages
## [7] # from Jan-1950 to Sep-2022
## [8] #
## [9] # WARNING: Values after Dec-2004 are updated
## [10] # using Satellite SSHa from AVISO Delayed Time product.
## [11] # http://opendap.aviso.oceanobs.com/thredds/dodsC/dataset-duacs-dt-global-allsat-msla
## [12] #
## [13] # PRELIMINARY: Values after Feb-2022 are preliminary and updated
## [14] # using Satellite SSHa from AVISO Near Real Time product.
## [15] # http://opendap.aviso.oceanobs.com/thredds/dodsC/dataset-duacs-nrt-over30d-global-al
## [16] #
## [17] # The update is performed by taking the NPGO spatial pattern of Di Lorenzo et al. 2008
## [18] # computed over the period 1950-2004, and projecting the AVISO Satellite SSHa.
## [19] # During the pre-processing of the AVISO data, we remove the seasonal cycle based on
## [20] # the 1993-2004 seasonal means.
## [21] #
## [22] # AVISO PRODUCT UPDATE Summer 2014: AVISO has released a re-processed dataset for the
## [23] # Starting from the November 2014, the NPGO index is computed with this updated dataset
## [24] # values from 2004 onward have been recomputed with very minor differences from previous
## [25] #
## [26] # Ref:
## [27] # Di Lorenzo et al., 2008: North Pacific Gyre Oscillation
## [28] # links ocean climate and ecosystem change, GRL.
## [29] #
## [30] #          YEAR          MONTH          NPGO index

```

Next, we extract the actual NPGO indices for the years of interest and inspect the file contents. We also want the average NPGO annual index from January 1 through December 31 during the first year that the juvenile steelhead are in the ocean (i.e., during their second year of life). Therefore, we need NPGO values from `yr_first + marine_lag == 1980` through `yr_last - age_min + n_fore + marine_lag == 2017`.

```

## number of years of data
n_yrs <- yr_last - yr_first + 1
## NPGO data for years of interest
dat_NPGO <- read_table(url_NPGO, col_names = FALSE,
                      skip = hdr_NPGO + (yr_first-1950)*12,
                      n_max = (n_yrs-1)*12)
colnames(dat_NPGO) <- c("year", "month", "NPGO")
## select only years of interest indexed by brood year
dat_NPGO <- dat_NPGO[dat_NPGO$year >= yr_first+marine_lag &
                    dat_NPGO$year <= yr_last-age_min+n_fore+marine_lag,]
dat_NPGO <- aggregate(dat_NPGO$NPGO, by = list(year = dat_NPGO$year), mean)
dat_NPGO <- data.frame(year = seq(yr_first, yr_last-age_min+n_fore),
                      NPGO = dat_NPGO[,2])
dat_NPGO[, "NPGO"] <- round(dat_NPGO[, "NPGO"], 2)

```

Hatchery releases

The numbers of hatchery fish released each year are provided in `/data/skagit_sthd_hrel.csv`. They have already been lagged 2 years (i.e., brood year + 2) to account for the potential competitive interactions during their juvenile life stage.

```
## get releases
dat_hrel <- read_csv(file.path(datadir, "skagit_sthd_hrel.csv"))
dat_hrel <- subset(dat_hrel, dat_hrel$year <= max(dat_NPGO$year))
```

Archive covariates

The last thing we will do is combine the covariates into one data frame and write them to a file for use in the analysis.

```
## combine covariates
dat_cvrs <- Reduce(function(...) merge(..., all = TRUE),
                    list(dat_flow_wtr,
                         dat_flow_sum,
                         dat_NPGO,
                         dat_hrel))
## check table of covariates
head(dat_cvrs)

##   year flow_wtr flow_sum NPGO   Hrel
## 1 1978   362.5    78.44 -0.75 340416
## 2 1979   594.7    66.26 -0.22 274090
## 3 1980   792.9    76.46 -0.69 216537
## 4 1981   538.0   106.19 -0.09 300942
## 5 1982   407.8   135.35  0.62 241931
## 6 1983   410.6   119.50 -0.39 256238

## write covariates to a file
write_csv(dat_cvrs, file.path(datadir, "skagit_sthd_covars.csv"))
```