R1. Instructions for retrieving and archiving the environmental covariates.

2024 - 2025 Skagit River steelhead forecast.

Contents

This is version 0.24.12.10.

1 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, they 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**.

```
if(!require("readr")) {
  install.packages("readr")
  library("readr")
}
if(!require("here")) {
  install.packages("here")
  library("here")
```

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

2 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_frst <- 1978
yr_last <- 2024

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

## time lags (years) for covariates
flow_lag <- 1
marine_lag <- 2
hrel_lag <- 2

## number of years for run forecasts
n_fore <- 1
```

3 Retrieve covariates

Our analysis investigates 3 covariates as possible drivers of the population's instrinic growth rate:

1. Maximum river discharge in winter;
2. Minimum river discharge in summer;
3. North Pacific Gyre Oscillation;

3.1 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 (#12200500), beginning with the first year of fish data.

```
## flow gage ID
flow_site <- 12200500
## get URL for flow data from USGS
flow_url <- paste0("https://waterdata.usgs.gov/nwis/dv",
                   "?cb_00060=on",
```

```
                    "&format=rdb",
                    "&site_no=",flow_site,
                    "&begin_date=",yr_frst,"-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://waterdata.usgs.gov/provisional-data-statement/
##  [9] #
## [10] # Contact:   gs-w_waterdata_support@usgs.gov
## [11] # retrieved: 2024-12-10 14:52:46 EST       (vaww01)
## [12] #
## [13] # Data for the following 1 site(s) are contained in this file
## [14] #    USGS 12200500 SKAGIT RIVER NEAR MOUNT VERNON, WA
## [15] # -----------------------------------------------------------------------------------
## [16] #
## [17] # Data provided for site 12200500
## [18] #            TS   parameter     statistic     Description
## [19] #        149429       00060     00003     Discharge, cubic feet per second (Mean)
## [20] #
## [21] # Data-value qualification codes included in this output:
## [22] #     A  Approved for publication -- Processing and review completed.
## [23] #     P  Provisional data subject to revision.
## [24] #     e  Value has been estimated.
## [25] #
```

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)
```

```r
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   `149429_00060_00003` `149429_00060_00003_cd`
##   <chr>        <int> <date>                    <dbl> <chr>
## 1 USGS      12200500 1978-01-01                16300 A
## 2 USGS      12200500 1978-01-02                15800 A
## 3 USGS      12200500 1978-01-03                16000 A
## 4 USGS      12200500 1978-01-04                19100 A
## 5 USGS      12200500 1978-01-05                21000 A
## 6 USGS      12200500 1978-01-06                21100 A
```

We only need the 3rd and 4th columns, which contain the date (`datetime`) and daily flow measurements (`149429_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".

```r
## 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")]
```

3.1.1 Winter maximum

We are interested in the maximum of the daily peak flows from October through March 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 1 year so they align with the juvenile life stage. Therefore, the flow time series will begin in 1978 and end in 2023.

```r
## 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
```

```r
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"
```

3.1.2 Summer minimum

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

```r
## summer flows in year t
flow_sum <- subset(dat_flow, (month>=6 & month<=9)
                   & year >= yr_frst+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"
```

3.2 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. We begin by downloading the raw NPGO data and viewing the metadata.

```r
## URL for NPGO data
url_NPGO <- "https://www.o3d.org/npgo/data/NPGO.txt"
## 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] #  Last update 16-Aug-2024 by E. Di Lorenzo
##  [2] #  NPGO index monthly averages
##  [3] #  from Jan-1950  to  Aug-2024
##  [4] #
##  [5] #  WARNING: Values after Dec-2004 are updated
```

```
##  [6] #  using Satellite SSHa from AVISO Delayed Time product.
##  [7] #  http://opendap.aviso.oceanobs.com/thredds/dodsC/dataset-duacs-dt-global-allsat-msla
##  [8] #
##  [9] #  PRELIMINARY: Values after Dec-2022 are preliminary and updated
## [10] #  using Satellite SSHa from AVISO Near Real Time product.
## [11] #  http://opendap.aviso.oceanobs.com/thredds/dodsC/dataset-duacs-nrt-over30d-global-al
## [12] #
## [13] #  The update is performed by taking the NPGO spatial pattern of Di Lorenzo et al. 200
## [14] #  computed over the period 1950-2004, and projecting the AVISO Satellite SSHa.
## [15] #  During the pre-processing of the AVISO data, we remove the seasonal cycle based on
## [16] #  the 1993-2004 seasonal means.
## [17] #
## [18] #  AVISO PRODUCT UPDATE Summer 2014: AVISO has released a re-processed dataset for the
## [19] #  Starting from the November 2014, the NPGO index is computed with this updated datas
## [20] #  values from 2004 onward have been recomputed with very minor differences from previ
## [21] #
## [22] #  Ref:
## [23] #  Di Lorenzo et al., 2008: North Pacific Gyre Oscillation
## [24] #  links ocean climate and ecosystem change, GRL.
## [25] #
## [26] #      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_frst + marine_lag == 1980` through `yr_last - age_min + n_fore + marine_lag == 2024`.

```
## number of years of data
n_yrs <- yr_last - yr_frst + 1
## NPGO data for years of interest
dat_NPGO <- read_table(url_NPGO, col_names = FALSE,
                       skip = hdr_NPGO + (yr_frst-1950)*12,
                       n_max = (n_yrs)*12-6)
colnames(dat_NPGO) <- c("year","month","NPGO")
## select only years of interest indexed by brood year
dat_NPGO <- dat_NPGO[dat_NPGO$year >= yr_frst+marine_lag &
                     dat_NPGO$year <= yr_last-age_min+n_fore+marine_lag,]

dat_NPGO[dat_NPGO$year >= yr_frst+marine_lag,]
```

```
## # A tibble: 534 x 3
##     year month    NPGO
##    <dbl> <dbl>   <dbl>
##  1  1980     1 -1.90
##  2  1980     2 -1.72
##  3  1980     3 -1.15
```

```
##   4  1980     4 -0.0658
##   5  1980     5 -0.211
##   6  1980     6 -0.683
##   7  1980     7 -0.990
##   8  1980     8 -1.08
##   9  1980     9 -0.697
##  10  1980    10 -0.0660
## # ... with 524 more rows
```

```
dat_NPGO <- aggregate(dat_NPGO$NPGO, by = list(year = dat_NPGO$year), mean)
dat_NPGO <- data.frame(year = seq(yr_frst,yr_last-age_min+n_fore),
                       NPGO = dat_NPGO[,2])
dat_NPGO[,"NPGO"] <- round(dat_NPGO[,"NPGO"], 2)
```

4 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
                       ))
## check table of covariates
head(dat_cvrs)
```

```
##   year flow_wtr flow_sum  NPGO
## 1 1978    988.3   146.96 -0.75
## 2 1979   2763.7   146.68 -0.22
## 3 1980   2860.0   147.81 -0.69
## 4 1981   1526.3   220.30 -0.09
## 5 1982   1789.6   222.57  0.62
## 6 1983   2302.2   158.57 -0.39
```

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