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

Retrospective analysis of Skagit River Chum salmon productivity

Contents

This is version 0.20.03.03.

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 <- 1980
yr_last <- 2018

## min & max adult ages (years)
age_min <- 3
age_max <- 5

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


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

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;
4. Hatchery SAR
5. Whidbey basin pink salmon escapement

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://help.waterdata.usgs.gov/policies/provisional-data-statement
##  [9] #
## [10] # File-format description:  https://help.waterdata.usgs.gov/faq/about-tab-delimited-ou
## [11] # Automated-retrieval info: https://help.waterdata.usgs.gov/faq/automated-retrievals
## [12] #
## [13] # Contact:   gs-w_support_nwisweb@usgs.gov
## [14] # retrieved: 2019-07-31 16:34:58 EDT       (vaww02)
## [15] #
## [16] # Data for the following 1 site(s) are contained in this file
## [17] #    USGS 12200500 SKAGIT RIVER NEAR MOUNT VERNON, WA
## [18] # -----------------------------------------------------------------------------------
## [19] #
## [20] # Data provided for site 12200500
## [21] #            TS   parameter     statistic     Description
## [22] #        149429       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 `149429_00060_00003` `149429_00060_00003_cd`
```

```
##       <chr>    <int>     <date>                  <dbl>                   <chr>
## 1      USGS 12200500 1980-01-01                16500                       A
## 2      USGS 12200500 1980-01-02                16700                       A
## 3      USGS 12200500 1980-01-03                17700                       A
## 4      USGS 12200500 1980-01-04                16400                       A
## 5      USGS 12200500 1980-01-05                15800                       A
## 6      USGS 12200500 1980-01-06                15500                       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".

```
## 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 November 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 1980 and end in 2016.

```
## autumn flows in year t
flow_aut <- subset(dat_flow, (month>=11 & 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<=3)
                   & 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), mean)
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"
print(dat_flow_wtr)
```

```
##    year flow_wtr
## 1  1980   2860.0
## 2  1981   1526.3
## 3  1982   1789.6
## 4  1983   2302.2
## 5  1984    775.9
## 6  1985   1795.3
## 7  1986   1843.4
## 8  1987    909.0
## 9  1988   1336.6
## 10 1989   2497.5
## 11 1990   4021.0
## 12 1991   1135.5
## 13 1992    781.5
## 14 1993   1030.7
## 15 1994   1577.2
## 16 1995   3737.8
## 17 1996   2089.8
## 18 1997   1008.1
## 19 1998   1469.6
## 20 1999   2174.7
## 21 2000    546.5
## 22 2001   2086.9
## 23 2002   1500.8
## 24 2003   1826.4
## 25 2004   1891.6
## 26 2005   1625.4
## 27 2006   3539.6
## 28 2007   2019.0
## 29 2008   2064.3
## 30 2009   1868.9
## 31 2010   2364.5
## 32 2011   1619.7
## 33 2012   1141.2
## 34 2013   1381.9
## 35 2014   2418.3
## 36 2015   2070.0
```

3.1.2 Spring max

Retrieving the flow juveniles would experience during their first spring and early summer rearing in Skagit Bay (April through June) is straightforward.

```
## spring flows in year t
flow_spr<- subset(dat_flow, (month>=3 & month<=6)
```

```r
                   & year >= yr_frst+flow_lag
                   & year <= yr_last-age_min+n_fore+flow_lag)
## change year index to brood year
flow_spr[,"year"] <- flow_spr[,"year"] - flow_lag
## combined flows indexed to brood year & calculate average flow
dat_flow_spr <- aggregate(flow ~ year, data = flow_spr, max)
dat_flow_spr <- round(dat_flow_spr, 2)
## for plotting purpose later
colnames(dat_flow_spr)[2] <- "flow_spr"
print(dat_flow_spr)
```

```
##    year flow_spr
## 1  1980   917.46
## 2  1981  1347.88
## 3  1982   846.67
## 4  1983   923.13
## 5  1984  1175.15
## 6  1985  1030.73
## 7  1986  1118.51
## 8  1987  1095.86
## 9  1988   846.67
## 10 1989   841.01
## 11 1990   869.33
## 12 1991  1160.99
## 13 1992  1022.24
## 14 1993  1030.73
## 15 1994   736.24
## 16 1995   906.14
## 17 1996  2089.78
## 18 1997   682.44
## 19 1998  1197.80
## 20 1999  1033.56
## 21 2000   744.73
## 22 2001  1452.65
## 23 2002  1279.92
## 24 2003   753.23
## 25 2004   569.17
## 26 2005  1095.86
## 27 2006  2101.11
## 28 2007  1735.82
## 29 2008   886.32
## 30 2009   843.84
## 31 2010  1177.98
## 32 2011  1628.22
## 33 2012  1325.23
## 34 2013  1381.86
## 35 2014   673.94
## 36 2015   761.72
```

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.

```
## 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 17-Jul-2019 by E. Di Lorenzo
##  [6] #  NPGO index monthly averages
##  [7] #  from Jan-1950  to  Jul-2019
##  [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 Jan-2019 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-all
## [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 datase
## [24] #  values from 2004 onward have been recomputed with very minor differences from previo
## [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_frst + marine_lag == 1981` through `yr_last - age_min + n_fore + marine_lag == 2016`.

```
## 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-1)*12)
colnames(dat_NPGO) <- c("year","month","NPGO")
## select only years of interest indexed by brood year
dat_NPGO_wtr <- subset(dat_NPGO, (month == 12)
                   & year >= yr_frst
                   & year <= yr_last-age_min+n_fore)

dat_NPGO_spr <- subset(dat_NPGO, (month >= 1 & month <= 3)
                   & year >= yr_frst+marine_lag
                   & year <= yr_last-age_min+n_fore+marine_lag)

## change spr year index to match wtr
dat_NPGO_spr[,"year"] <- dat_NPGO_spr[,"year"] - marine_lag

## combined NPGO indexed to brood year & calculate december - March average
dat_NPGO <- aggregate(NPGO~year, data = rbind(dat_NPGO_wtr,dat_NPGO_spr), 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)
```

3.3 Hatchery Smolt to adult survival rates

We used a time series of marine survival of hatchery Chum salmon from the Tulalip Hatchery as an indicator of marine survival for conspecific wild Chum salmon from the Skagit River.

```
dat_SAR <- read_csv(file.path(datadir,"ps_hatchery_chum_return_rates.csv"))
dat_SAR <- subset(dat_SAR, year >= yr_frst
                   & year <= yr_last-age_min+n_fore)

dat_SAR <- data.frame(dat_SAR)

dat_SAR <- dat_SAR[,c(1,4)]
print(dat_SAR)
```

```
##    year    Tulalip
## 1  1980  3.3879987
## 2  1981  5.6751522
## 3  1982 14.4417349
## 4  1983  2.1223123
## 5  1984 11.0153670
```

```
## 6   1985   8.7570109
## 7   1986 10.4798662
## 8   1987   4.0466845
## 9   1988   3.0621649
## 10  1989   2.2609455
## 11  1990   9.9789032
## 12  1991   1.3910264
## 13  1992   3.6152615
## 14  1993   2.2592288
## 15  1994   2.2719406
## 16  1995   4.4830040
## 17  1996   1.1566535
## 18  1997   5.3983225
## 19  1998 14.6207124
## 20  1999   6.7131602
## 21  2000   3.7462885
## 22  2001   5.1323788
## 23  2002   2.6584741
## 24  2003   2.7421580
## 25  2004   6.6129158
## 26  2005   3.7972014
## 27  2006   2.0014710
## 28  2007   3.5351005
## 29  2008   0.6264500
## 30  2009   4.0434074
## 31  2010   1.7306812
## 32  2011   0.3638305
```

3.4 Pink salmon escapement

We used pink salmon escapement as a covariate in the model. The plausible hypothesis is that pink salmon fry compete with Chum fry for food and space and therefore may reduce the overall prodcutivity of Chum salmon.

```
dat_pink_esc <- read_csv(file.path(datadir,"skagit_pink_esc.csv"))
dat_pink_esc <- subset(dat_pink_esc, year >= yr_frst
                   & year <= yr_last-age_min+n_fore)
dat_pink_esc <- data.frame(dat_pink_esc[,c(1,3)])
```

```
print(dat_pink_esc)
```

```
##    year whidbey_basin_pink_escapement
## 1  1980                             0
## 2  1981                        208728
## 3  1982                             0
## 4  1983                        794922
## 5  1984                             0
## 6  1985                       1212444
```

```
## 7  1986                              0
## 8  1987                         866906
## 9  1988                              0
## 10 1989                         551870
## 11 1990                              0
## 12 1991                         611447
## 13 1992                              0
## 14 1993                         740135
## 15 1994                              0
## 16 1995                        1166626
## 17 1996                              0
## 18 1997                         252109
## 19 1998                              0
## 20 1999                         781543
## 21 2000                              0
## 22 2001                        2741709
## 23 2002                              0
## 24 2003                        2144081
## 25 2004                              0
## 26 2005                         660124
## 27 2006                              0
## 28 2007                        1683591
## 29 2008                              0
## 30 2009                        3992373
## 31 2010                              0
## 32 2011                        1172903
## 33 2012                              0
## 34 2013                        3053569
## 35 2014                              0
## 36 2015                         770674
```

```
dat_pink_esc <- dat_pink_esc[which(dat_pink_esc$year %in% seq(yr_frst,yr_last,1)),]
print(dat_pink_esc)
```

```
##    year whidbey_basin_pink_escapement
## 1  1980                              0
## 2  1981                         208728
## 3  1982                              0
## 4  1983                         794922
## 5  1984                              0
## 6  1985                        1212444
## 7  1986                              0
## 8  1987                         866906
## 9  1988                              0
## 10 1989                         551870
## 11 1990                              0
## 12 1991                         611447
## 13 1992                              0
## 14 1993                         740135
```

```
## 15 1994                          0
## 16 1995                    1166626
## 17 1996                          0
## 18 1997                     252109
## 19 1998                          0
## 20 1999                     781543
## 21 2000                          0
## 22 2001                    2741709
## 23 2002                          0
## 24 2003                    2144081
## 25 2004                          0
## 26 2005                     660124
## 27 2006                          0
## 28 2007                    1683591
## 29 2008                          0
## 30 2009                    3992373
## 31 2010                          0
## 32 2011                    1172903
## 33 2012                          0
## 34 2013                    3053569
## 35 2014                          0
## 36 2015                     770674
```

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.

```
#print(dat_flow_wtr)
#print(dat_flow_spr)
## combine covariates
dat_cvrs <- Reduce(function(...) merge(..., all = TRUE),
                   list(dat_flow_wtr,
                        dat_flow_spr,dat_NPGO,
                        dat_pink_esc))

## check table of covariates
print(dat_cvrs)
```

```
##    year flow_wtr flow_spr  NPGO whidbey_basin_pink_escapement
## 1  1980   2860.0   917.46 -0.57                             0
## 2  1981   1526.3  1347.88 -0.29                        208728
## 3  1982   1789.6   846.67  0.32                             0
## 4  1983   2302.2   923.13  0.06                        794922
## 5  1984    775.9  1175.15  0.19                             0
## 6  1985   1795.3  1030.73 -1.09                       1212444
## 7  1986   1843.4  1118.51  0.32                             0
```

```
## 8  1987   909.0  1095.86  0.89                        866906
## 9  1988  1336.6   846.67  0.91                             0
## 10 1989  2497.5   841.01  0.12                        551870
## 11 1990  4021.0   869.33 -0.85                             0
## 12 1991  1135.5  1160.99 -0.25                        611447
## 13 1992   781.5  1022.24 -2.14                             0
## 14 1993  1030.7  1030.73 -1.92                        740135
## 15 1994  1577.2   736.24 -1.28                             0
## 16 1995  3737.8   906.14 -0.66                       1166626
## 17 1996  2089.8  2089.78 -1.18                             0
## 18 1997  1008.1   682.44  0.90                        252109
## 19 1998  1469.6  1197.80  1.53                             0
## 20 1999  2174.7  1033.56  1.68                        781543
## 21 2000   546.5   744.73  2.59                             0
## 22 2001  2086.9  1452.65  1.99                       2741709
## 23 2002  1500.8  1279.92  1.77                             0
## 24 2003  1826.4   753.23  0.38                       2144081
## 25 2004  1891.6   569.17 -1.41                             0
## 26 2005  1625.4  1095.86 -0.92                        660124
## 27 2006  3539.6  2101.11 -0.31                             0
## 28 2007  2019.0  1735.82  0.88                       1683591
## 29 2008  2064.3   886.32  0.60                             0
## 30 2009  1868.9   843.84  1.66                       3992373
## 31 2010  2364.5  1177.98  0.71                             0
## 32 2011  1619.7  1628.22  0.74                       1172903
## 33 2012  1141.2  1325.23  1.07                             0
## 34 2013  1381.9  1381.86 -0.64                       3053569
## 35 2014  2418.3   673.94 -0.89                             0
## 36 2015  2070.0   761.72  0.13                        770674
```

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