Generating ASMR estimates for North Area BGC units

Hardy Griesbauer

17/09/2019

## Description of choosing climate stations for BGC units in the North Area

Below is R code and description used to select climate station data to generate ASMR estimates for various BGC units in the North Area.

# Load libraries  
library(forestDroughtTool)  
library(tidyverse)  
library(weathercan)  
library(ggplot2)  
library(here)  
library(magrittr)

## What are the BGC units of interest in the North Area?

What units do we want to generate drought hazard estimates for?

# North Area BGC units are saved in .rda file  
load(here("dat","northBGC.rda"))  
northBGC

## [1] "ICHmc1" "MHmm2" "SBSmc2" "CWHws2" "ICHmc2" "ESSFmv3" "ESSFmv1"  
## [8] "SBSwk3" "SBSdw3" "SBSwk2" "ESSFmv2" "ESSFmv4" "ESSFwc3" "ESSFwk2"  
## [15] "BWBSwk2" "SBSmk1" "SBSmk2" "SBSwk1" "SBSmc3" "ESSFxv1" "SBSdw1"   
## [22] "SBSdw2" "ICHvk2" "ICHwk3" "ICHwk4" "ESSFwk1" "ESSFmm1" "IDFdk3"   
## [29] "SBSmc1"

## What ECCC climate stations exist for these units?

Using a bunch of different packages, we can query the ECCC climate station data and see which ones are located in the BGC units of interest.

# DON'T RUN - Takes a long time, instead use load()  
# # Step 1 - Select stations in BC with daily data and assign to 'stn'  
# stn<-  
# weathercan::stations %>%   
# dplyr::filter(prov=="BC" & interval=="day") %>%   
#   
# # Step 2 - Convert to a spatial file and merge with BGC units (this will take awhile!)  
# sf::st\_as\_sf(coords=c("lon","lat")) %>% # convert to spatial file  
# sf::st\_set\_crs(4326) %>% # set to WGS1984 datum  
# bcmaps::transform\_bc\_albers() %>% # set to BC albers projection  
# sf::st\_join(bcmaps::bec()[,"MAP\_LABEL"]) %>% # merge with BEC  
# sf::st\_drop\_geometry() %>% # drop geometry  
#   
# # Step 3 - Merge with BGC units of interest (above)  
# dplyr::filter(MAP\_LABEL%in%northBGC) %>% # filter for BGC units of interest  
# dplyr::mutate(length=end-start+1) %>% # create a column to show record length  
# dplyr::arrange(MAP\_LABEL,desc(length)) %>%   
# dplyr::select(station\_name,station\_id,bgc=MAP\_LABEL,start,end,length)  
#   
# # step 4 save stn so you don't have to run above lines of code  
# save(stn,file=here("dat","stationList.rda"))  
  
# Load station data compiled from above lines of code  
 load(here("dat","stationList.rda"))  
  
# Print to screen  
 stn

## # A tibble: 154 x 6  
## station\_name station\_id bgc start end length  
## <chr> <fct> <chr> <int> <int> <dbl>  
## 1 PINK MOUNTAIN 1432 BWBSwk2 1973 1982 10  
## 2 PINK MOUNTAIN 2 1430 BWBSwk2 1982 1990 9  
## 3 BUCKINGHORSE RIVER 1450 BWBSwk2 1973 1977 5  
## 4 TAHTSA LAKE WEST 547 CWHws2 1951 2000 50  
## 5 ALICE ARM 375 CWHws2 1948 1964 17  
## 6 KITSAULT MINESITE 474 CWHws2 1969 1972 4  
## 7 KITSAULT MINESITE 405 CWHws2 1968 1969 2  
## 8 CRANBERRY RIVER 460 CWHws2 1973 1974 2  
## 9 NASS RIVER 483 CWHws2 1956 1956 1  
## 10 FIRVALE 6941 CWHws2 1991 1991 1  
## # ... with 144 more rows

## Download climate data

Need to download climate data for stations. This will take a long time, so I saved the downloaded data into an .rda file.

* Note that in the future we will want to update some of the station data with cleaned data from Vanessa. This will take me some time!\*

I printed out the station list, and manually selected stations (not shown) that have long records that span the 1961-1990 climate normal period, as much as possible. In some cases, we may want to join records.

# Row numbers of stations of interest  
stnRow=c(1, 2, 4, 11, 12, 15, 17, 21, 22, 36, 37, 39, 66, 67,  
 72, 73, 85, 86, 87, 93, 108, 109, 123, 137, 143,  
 144, 149, 150, 151, 153)  
  
# Don't run - takes a long time. Use load() instead  
# Download data using the stnRow (this can take a long time!)  
# climData<-  
# weather\_dl(station\_ids=stn$station\_id[stnRow],interval="day") %>%  
#   
# # format climate data (see below for more information)  
# select(stn=station\_name,date,tmn=min\_temp,tmx=max\_temp,ppt=total\_precip,year,month,day)  
  
# save(climData,file=here("dat","climData.rda"))  
  
load(here("dat","climData.rda"))

## Clean climate station data

We need to process missing values in the daily climate data. For this project, we: - omitted any years with >10 consecutive missing data in any climate variable;and - imputed missing data using adjacent values (closest data before and after missing value) - Leap years have 366 days, so we removed February 29 from those years to keep number of days to 365.

# DON'T RUN - this takes a long time, use load() instead  
  
# cleanECData() for each station  
# x1<-by(INDICES=climData$stn,function(x) cleanECData(x),data=climData)   
#   
# # Formatting  
# climData\_cleaned<-dplyr::bind\_rows(x1,.id="id")   
# climData\_cleaned$stn<-names(x1)[as.numeric(climData\_cleaned$id)]  
# rm(x1)  
  
# # Format date columns  
# climData\_cleaned%<>%  
# mutate(month=as.integer(month)) %>%   
# mutate(day=as.integer(day)) %>%   
# mutate(year=as.integer(year)) %>%   
# mutate(date=paste(year,month,day,sep="-"))  
  
# Rename columns to make it easier to pass to asmrCalc()  
# climData\_cleaned%<>%  
# rename(!!'tmn':=tmn\_filled,!!'ppt':=ppt\_filled,!!'tmx':=tmx\_filled)  
  
# save(climData\_cleaned,file=here("dat","climData\_cleaned.rda"))  
  
# Load cleaned EC climate data from above lines of code  
load(here("dat","climData\_cleaned.rda"))  
  
# Summarize station data   
climData\_cleaned %>%  
 group\_by(stn,year) %>%  
 summarise(n()) %>%  
 ungroup() %>%  
 group\_by(stn) %>%  
 summarise(Num.years=n(),start=min(year),end=max(year))

## Warning: Grouping rowwise data frame strips rowwise nature

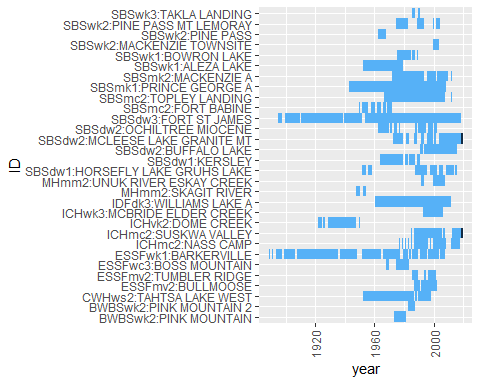
## # A tibble: 30 x 4  
## stn Num.years start end  
## <chr> <int> <int> <int>  
## 1 ALEZA LAKE 27 1953 1979  
## 2 BARKERVILLE 102 1889 2007  
## 3 BOSS MOUNTAIN 11 1968 1983  
## 4 BOWRON LAKE 12 1976 1989  
## 5 BUFFALO LAKE 24 1991 2015  
## 6 BULLMOOSE 15 1987 2002  
## 7 DOME CREEK 24 1922 1950  
## 8 FORT BABINE 16 1950 1971  
## 9 FORT ST JAMES 118 1895 2018  
## 10 HORSEFLY LAKE GRUHS LAKE 28 1952 2015  
## # ... with 20 more rows

## Figure showing data coverage for each station

It might be useful to produce a figure showing climate data coverage for each station and BGC unit.

climData\_cleaned %>%   
 group\_by(stn,year) %>%   
 summarise(days=n()) %>%   
 inner\_join(stn,by=c("stn"="station\_name")) %>%   
 select(stn,year,days,bgc) %>%   
 ungroup() %>%   
 mutate(ID=paste(bgc,stn,sep=":")) %>%   
 ggplot(aes(year, ID)) +  
 geom\_tile(aes(fill = days))+  
 theme(legend.position = "none")+  
 theme(axis.text.x = element\_text(angle = 90,vjust = 0.5))

## Warning: Grouping rowwise data frame strips rowwise nature



## Select stations and climate data to run model

Generally, we have been summarising site ASMR using 10 years of daily climate data. The rationale for this is that we want to keep data consistent between BGC units, and we need to be able to use stations with relatively sparse data (ie. 10 years).

Based on a conversation with Vanessa, we’ve decided on the following:

1. Restrict data to 1940 and 1990. Data pre-1940 may have measurement errors, and post-1990 is starting to get into the “climate change” realm.
2. Ideally, restrict years to 1961-1990 normal period. If number of years available is less than 10, then use years from 1940-1990.
3. Select 10 years randomly from within the dataset. If there are less than 10 years of data within 1961-1990 dataset, then use all the years from 1961-1990, and add random years’ data from 1940-1960 until number of years = 10.

This is all taken care of in the asmrSelect() function in the forestDroughtTool package.

# Need to change Pink Mountain2 to Pink Mountain, to concatenate the datasets  
climData\_cleaned %<>%   
 mutate(stn=replace(stn,stn=="PINK MOUNTAIN 2","PINK MOUNTAIN"))  
  
# Designate stations to use:  
stnList=unique(climData\_cleaned$stn)[c(2,3,4,6,9,11,12,16,17,19,20,21,24,26,29)]  
stnID<-stn[stn$station\_name%in%stnList,]  
stnID<-stnID[stnID$station\_id!="50820",] # get rid of extra Williams Lake data  
  
# Print out stations used to estimate current drought hazard in North Area  
stnID

## # A tibble: 15 x 6  
## station\_name station\_id bgc start end length  
## <chr> <fct> <chr> <int> <int> <dbl>  
## 1 PINK MOUNTAIN 1432 BWBSwk2 1973 1982 10  
## 2 TAHTSA LAKE WEST 547 CWHws2 1951 2000 50  
## 3 BULLMOOSE 1397 ESSFmv2 1982 2003 22  
## 4 BOSS MOUNTAIN 569 ESSFwc3 1967 1984 18  
## 5 BARKERVILLE 568 ESSFwk1 1888 2015 128  
## 6 NASS CAMP 482 ICHmc2 1973 2017 45  
## 7 WILLIAMS LAKE A 664 IDFdk3 1960 2012 53  
## 8 KERSLEY 601 SBSdw1 1962 1992 31  
## 9 OCHILTREE MIOCENE 623 SBSdw2 1963 2003 41  
## 10 FORT ST JAMES 588 SBSdw3 1895 2018 124  
## 11 TOPLEY LANDING 496 SBSmc2 1962 2017 56  
## 12 PRINCE GEORGE A 631 SBSmk1 1942 2009 68  
## 13 MACKENZIE A 1423 SBSmk2 1971 2013 43  
## 14 BOWRON LAKE 564 SBSwk1 1974 1990 17  
## 15 PINE PASS MT LEMORAY 1429 SBSwk2 1974 2004 31

Now let’s run a loop to generate the dataset and export it.

# Don't run this takes a long time, use load () instead  
# for (j in 1:nrow(stnID)) {  
#   
# # Get ASMR summaries for the station  
# x<-  
# climData\_cleaned %>%   
# filter(stn==stnID$station\_name[j]) %>%   
# asmrCalc() %>%   
# asmrSelect()   
#   
# # Extract ASMR mean and SD  
# asmr<-  
# x$asmr %>%   
# mutate(bgc=stnID$bgc[j]) %>%   
# dplyr::select(bgc,month,everything())  
#   
# # Extract years used to calculate ASMR  
# yearASMR<-  
# c(stnID$bgc[j],x$years)  
#   
# # Compile asmr into data frame  
# if(j==1) {asmrNorth=asmr} else {asmrNorth=rbind(asmrNorth,asmr)}   
#   
# # Compile years into data frame  
# if(j==1) {years=yearASMR} else {years=rbind(years,yearASMR)}   
#   
# if (j==nrow(stnID)) {  
#   
# years=as.data.frame(years)  
# names(years)=c("bgc",paste("year",1:10,sep=""))  
# years<-as\_tibble(years)  
#   
# }  
#   
# }  
   
 # save(asmrNorth,file="asmrNorth.rda")  
 # save(years,file="years.rda")  
   
load(here::here("dat","asmrNorth.rda"))

# ANALYSIS

## Generate annual ASMR estimates by BGC unit

# Generate annual AET/PET for five sites by BGC unit  
asmrANNUAL<-  
 asmrNorth %>%   
 dplyr::select(bgc,month,contains(".mean")) %>%   
 group\_by(bgc) %>%   
 summarise\_at(vars(contains(".mean")),mean)   
  
knitr::kable(asmrANNUAL,digits=2)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| bgc | S1.mean | S2.mean | S3.mean | S4.mean | S5.mean |
| BWBSwk2 | 0.84 | 0.92 | 0.97 | 1.00 | 1 |
| CWHws2 | 1.00 | 1.00 | 1.00 | 1.00 | 1 |
| ESSFmv2 | 0.90 | 0.95 | 0.97 | 0.99 | 1 |
| ESSFwc3 | 1.00 | 1.00 | 1.00 | 1.00 | 1 |
| ESSFwk1 | 0.98 | 0.99 | 1.00 | 1.00 | 1 |
| ICHmc2 | 0.86 | 0.90 | 0.93 | 0.96 | 1 |
| IDFdk3 | 0.70 | 0.76 | 0.79 | 0.84 | 1 |
| SBSdw1 | 0.72 | 0.79 | 0.82 | 0.87 | 1 |
| SBSdw2 | 0.80 | 0.85 | 0.88 | 0.92 | 1 |
| SBSdw3 | 0.78 | 0.83 | 0.85 | 0.90 | 1 |
| SBSmc2 | 0.82 | 0.87 | 0.89 | 0.94 | 1 |
| SBSmk1 | 0.82 | 0.88 | 0.92 | 0.96 | 1 |
| SBSmk2 | 0.85 | 0.90 | 0.93 | 0.97 | 1 |
| SBSwk1 | 0.89 | 0.95 | 0.98 | 1.00 | 1 |
| SBSwk2 | 0.86 | 0.92 | 0.96 | 0.97 | 1 |

# write.csv(asmrANNUAL,file="asmrANNUAL.csv")  
  
# Classify  
asmrCut<-function(x) cut(x,breaks=c(0,asmrClass$asmrUL),labels=asmrClass$class)  
  
asmrANNUAL %>%   
 mutate\_if(is.numeric,asmrCut) %>%   
 knitr::kable()

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| bgc | S1.mean | S2.mean | S3.mean | S4.mean | S5.mean |
| BWBSwk2 | MD | SD | F | F | F |
| CWHws2 | F | F | F | F | F |
| ESSFmv2 | SD | SD | F | F | F |
| ESSFwc3 | F | F | F | F | F |
| ESSFwk1 | F | F | F | F | F |
| ICHmc2 | SD | SD | SD | F | F |
| IDFdk3 | VD2 | MD | MD | MD | F |
| SBSdw1 | VD2 | MD | MD | SD | F |
| SBSdw2 | MD | SD | SD | SD | F |
| SBSdw3 | MD | MD | SD | SD | F |
| SBSmc2 | MD | SD | SD | SD | F |
| SBSmk1 | MD | SD | SD | F | F |
| SBSmk2 | SD | SD | SD | F | F |
| SBSwk1 | SD | F | F | F | F |
| SBSwk2 | SD | SD | F | F | F |