Getting Started with ACLIM ROMSNPZ data

Kirstin Holsman

Contents

Figure 1: .

ACLIM Repo: github.com/kholsman/ACLIM_romsnpz   and data files maintained by:
Kirstin Holsman
Alaska Fisheries Science Center
NOAA Fisheries, Seattle WA
kirstin.holsman@noaa.gov
Last updated: Jan 20, 2021

1. Overview

This is an overview of ACLIM plotting code and "canned" Rdata files generated from the downscaled ROM-SNPZ modeling work ACLIM modelers Drs. Hermann, Cheng, Kearney,Pilcher, and Aydin. Dr. Kelly

Kearney has recently dedicated significant time and energy towards organizing and documenting the ROM-SNPZ output, especially as it pertains to the ACLIM project. We strongly recommend reviewing this documentation before using the data in order to understand the origin of the indices and their present level of skill and validation, which varies considerably across indices and in space and time.

The following reading is recommended:

1.1. Downscaled models and carbon scenarios

The full ACLIM "suite" of models include are summarized in the following table:

Table 1: Summary of downscaled model runs based on boundary conditions forced by General Circulation Models (GCM) run under Coupled Model Intercomparison Project (CMIP) phase 5 (5th IPCCAssessment Report) or phase 6 (6th IPCC Assessment Report; "AR") global carbon mitigation scenarios. For full details see the Kearney 2021 Tech. Memo.

CMIP	GCM	Scenario	Def	Years	Model	Status	Source
5	GFDL	RCP 4.5	Med. mitigation	2006 - 2099	H16	ACLIM/FATE	Public
5	GFDL	RCP 8.5	High baseline	2006 - 2099	H16	ACLIM/FATE	Public
5	GFDL	RCP 8.5bio*	High baseline	2006 - 2099	H16	ACLIM/FATE	Public
5	MIROC	RCP 4.5	Med. mitigation	2006 - 2099	H16	ACLIM/FATE	Public
5	MIROC	RCP 8.5	High baseline	2006 - 2099	H16	ACLIM/FATE	Public
5	CESM	RCP 4.5	Med. mitigation	2006 - 2099	H16	ACLIM/FATE	Public
5	CESM	RCP 8.5	High baseline	2006 - 2080	H16	ACLIM/FATE	Public
5	CESM	RCP 8.5bio*	High baseline	2006 - 2099	H16	ACLIM/FATE	Public
	CORECFS	Reanalysis	Hindcast	1970 - 2018	H16	ACLIM	Public
	CORECFS	Reanalysis	Hindcast	1970 - 2020	K20	ACLIM2/RTAP	Public
6	CESM	SSP585	High baseline	2014 - 2099	K20P19	ACLIM2/RTAP	Embargo
6	CESM	SSP126	High Mitigation	2014 - 2099	K20P19	ACLIM2/RTAP	Embargo
6	CESM	Historical	Historical	1980 - 2014	K20P19	ACLIM2/RTAP	Embargo
6	GFDL	SSP585	High baseline	2014 - 2099	K20P19	ACLIM2/RTAP	Embargo
6	GFDL	SSP126	High Mitigation	2014 - 2099	K20P19	ACLIM2/RTAP	Embargo
6	GFDL	Historical	Historical	1980 - 2014	K20P19	ACLIM2/RTAP	Embargo
6	MIROC	SSP585	High baseline	2014 - 2099	K20P19	ACLIM2/RTAP	Embargo
6	MIROC	SSP126	High Mitigation	2014 - 2099	K20P19	ACLIM2/RTAP	Embargo
6	MIROC	Historical	Historical	1980 - 2014	K20P19	ACLIM2/RTAP	Embargo

*"bio" = nutrient forcing on boundary conditions

1.2. More information on the BERING10K ROMSNPZ model

1.2.1. The Bering 10K Model (v. H16) with 10 depth layers:

The H16 model is the original BSIERP era 10 depth layer model with a 10 Km grid. This version was used in ACLIM1.0 to dynamically downscaled 3 global circulation models (GCMs) under two CMIP5 representative carbon pathways (RCP): RCP 4.5 or "moderate global carbon mitigation" and RCP 8.5 "high baseline global carbon emissions". Details of the model and projections can be found in:

Hindcast (1979-2012; updated to 2016 during ACLIM 1.0): Hermann, A. J., G. A. Gibson, N. A. Bond, E. N. Curchitser, K. Hedstrom, W. Cheng, M. Wang, E. D. Cokelet, P. J. Stabeno, and K. Aydin.

2016. Projected future biophysical states of the Bering Sea. Deep Sea Research Part II: Topical Studies in Oceanography 134:30–47.doi:10.1016/j.dsr2.2015.11.001

Projections of the H16 10 layer model using CMIP5 scenarios: Hermann, A. J., G. A. Gibson, W. Cheng, I. Ortiz, K. Aydin, M. Wang, A. B. Hollowed, K. K. Holsman, and S. Sathyendranath. 2019. Projected biophysical conditions of the Bering Sea to 2100 under multiple emission scenarios. ICES Journal of Marine Science 76:1280–1304.doi:10.1093/icesjms/fsz043)

1.2.2. The Bering 10K Model (v. K20) with 30 depth layers and other advancements:

The Bering10K model was subsequently updated by Kearney et al. 2020 (30 layer and other NPZ updates) and Pilcher et al.2019 (OA and O2 dynamics) and this version is used for the projections in ACLIM2.0 under CMIP6.

Projections of the K20 30 layer model using CMIP6 scenarios: Hermann et al. in prep Cheng et al. in prep Kearney et al. in prep Pilcher et al. in prep (CMIP5 K20 projections) (ACLIM indices avail by permission only)

1.3. ACLIM ROMSNPZ outputs: Level3 ACLIM Indices (google drive)

1.3.1. Step 1: Clone the ACLIM_romsnpz repo to your local directory:

First clone the ACLIM ROMSNPZ repo. This repo will open and explore the netcdf (.nc) files in R and produce plots and standardized outputs for ACLIM analyses. Some standardized tools are included as functions in this repo including spatial averaging for seasonal, monthly and annual indices (e.g., Fall zooplankton biomass), as well as bias correction for projections (see Holsman et al. 2020 and Reum et al. 2020 for ACLIM 1.0 bias correction methods). The repo also includes a Rshiny interactive exploratory graphing tool which can be viewed online **at this link**.

Select `clone` or downlaod from the upper right hand side of the repo page :**github.com/kholsman/ACLIM_romsnpz** and save it to your local directory: `~[YOURPATH]/ACLIM_romsnpz`.
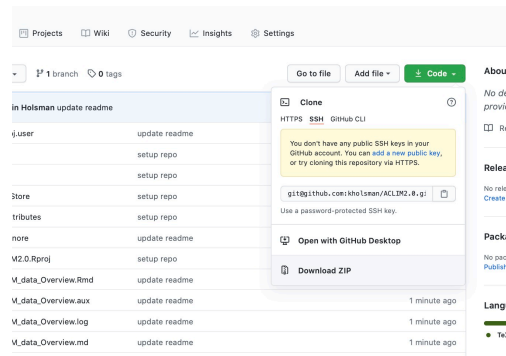
1.3.2 Step 2: Download the data from ACLIM google drive:

Data files are too large to store in the GitHub repository and are instead saved in the shared ACLIM data folder. For most applications you can use the ACLIM level3 post-processed indices available on the shared ACLIM drive: **00_ACLIM_shared>02_DATA>2021_01_19**.

Please be sure to coordinate with ROMSNPZ modeling team members to ensure data is applied appropriately. If you need access to the raw ROMSNPZ files (netcdf, non-regridded large files located on MOX). Please contact **Kelly Kearney** or **Al Hermann**.

IMPORTANT The ACLIM indices are stored as netcdf files (.nc) format in the Data folder of the ACLIM shared google drive. Please note that while the CMIP5 set is now public (Hermann et al. 2019) **the CMIP6 suite is under embargo for QAQC and should not be shared outside of the ACLIM group**. Al, Wei, Kelly, Darren, and Kerim are in the process of synthesizing and publishing the CMIP6 data (goal is spring 2021 for submission), following those publications the data will be made accessible to the public via the PMEL data portal, as is the case for the CMIP5 data and public hindcasts. It is strongly recommended that you include at least one (ideally multiple) authors from the ROMSNPZ team as co-author on your paper if you are linking to this data, this is especially the case for the CMIP6 data. There are multiple spatial and temporal caveats that are best described in discussions with the authors of these data and inclusion as co-authors will facilitate appropriate application and interpretation of the ROMSNPZ data.

These files can be used to generate seasonal, monthly, and annual indices (like those reported in Reum et al. 2020, Holsman et al. 2020).

The ACLIM Level3 indices are stored in a folder by YYYY_MM_DD (e.g.,`2021_01_19`) downloaded from the root google drive folder **00_ACLIM_shared>02_DATA**. The contents of the file (`roms_for_aclim` folder) should be copied into the `~[YOURPATH]/ACLIM2/Data/in/Newest` folder on your local repo directory, e.g.`~ACLIM2/Data/in/Newest/roms_for_aclim`. This folder is organized by Bering10K version, General Circulation Model (GCM) and carbon scenario, e.g. `B10K-H16_CMIP5_CESM_rcp45`. Within each folder the following subfolders are:

- `Level1`: (Empty; not copied from Mox)
- `Level2`: (Empty; not copied from Mox)
- `Level3`: 2 files (`ACLIMsurveyrep_B10K-x.nc` and `ACLIMregion_B10K-x.nc`)

1) `ACLIMsurveyrep_B10K-x.nc` contains summer groundfish trawl "survey replicated" indices (using mean date and lat lon) *(Note that the resampling stations need to be removed before creating bottom temperature maps)*

2) `ACLIMregion_B10K-x.nc`: contains weekly "strata" values *(Note that area weighting should be used to combine values across multiple strata)*

In section 2 below we explore these indices in more detail using R, including using (2) above to generate weekly, monthly, and seasonal indices (e.g. Fall Zooplankton) for use in biological models.

1.4. (Coming soon) PMEL public web-based database (beta-testing phase; currently limited to CMIP5)

The ROMSNPZ team has been working with Roland Schweitzer and Peggy Sullivan to develop the ACLIM Live Access Server (LAS) to publicly host the published CMIP5 hindcasts and downscaled projections. This server is in beta testing phase and can be accessed at the following links:

LAS custom ROMSNPZ data query, mapping, and plotting tool

THREDDS ACLIM data access tool

ERDAPP ACLIM data access tool

2. Exploring the ACLIM indices

The following examples show how to access and plot the ACLIM indices from their stored netcdf (.nc) format in the Data folder of the ACLIM shared google drive. Please note that while the CMIP5 set is now public (Hermann et al. 2019) **the CMIP6 suite is under embargo for QAQC and should not be shared outside of the ACLIM group**. Al, Wei, Kelly, Darren, and Kerim are in the process of synthesizing and publishing the CMIP6 data (goal is spring 2021 for submission), following those publications the data will be made accessible to the public via the PMEL data portal, as is the case for the CMIP5 data and public hindcasts. It is strongly recommended that you include at least one (ideally multiple) authors from the ROMSNPZ team as co-author on your paper if you are linking to this data, this is especially the case for the CMIP6 data. There are multiple spatial and temporal caveats that are best described in discussions with the authors of these data and inclusion as co-authors will facilitate appropriate application and interpretation of the ROMSNPZ data.

2.1. Explore the indices using ncdf4 and ggplot

The naming convention of the folders is: `B10K-[ROMSNPZ version]_[CMIP]_[GCM]_[carbon scenario]`.For example, the CMIP5 set of indices was downscaled using the H16 (Hermann et al. 2016) version of the ROMSNPZ. Three models were used to force boundary conditions(MIROC, CESM, and GFDL) under 2 carbon scenarios RCP 8.5 and RCP 4.5. So to see an individual trajectory we might look in the level3 (timeseries indices) folder under `B10K-H16_CMIP5_CESM_rcp45`, which would be the B10K version H16 of the CMIP5 CESM model under RCP4.5.

2.1.1 Set up the R Workspace and explore the two ACLIM Level3 indices types:

The following `make.R` script will load the directory paths, preferences, packages, and based functions into R().

```
tmstp   <- format(Sys.time(), "%Y_%m_%d")
main    <- getwd()  #"~/GitHub_new/ACLIM_romsnpz"

# loads packages, data, setup, etc.
source("R/make.R")
```

Once the base files and setup are loaded you can explore the index types. Recall that in each scenario folder there are two indices saved within the `Level3` subfolders:

1) `ACLIMsurveyrep_B10K-x.nc` contains summer groundfish trawl "survey replicated" indices (using mean date and lat lon) *(Note that the resampling stations need to be removed before creating bottom temperature maps)*

2) `ACLIMregion_B10K-x.nc`: contains weekly "strata" values *(Note that area weighting should be used to combine values across multiple strata)*

```
# list of the scenario x GCM downscaled ACLIM indices
for(k in aclim)
 cat(paste(k,"\n"))
```

```
# get some info about a scenario:

all_info1 <- info(model_list=aclim,type=1)
all_info2 <- info(model_list=aclim,type=2)
all_info1
all_info2

# variables in each of the two files:
srvy_vars
weekly_vars

#summary tables for variables
srvy_var_def
weekly_var_def
```

2.2. Explore "survey replicated" .nc files in R()

With the workspace setup the following code can be used to explore the ACLIM indices.The first indices we will explore are the survey replicated values for each variable. These are stored in the ACLIMsurveyrep_B10K-[version_CMIPx_GCM_RCP].nc in each scenario folder.
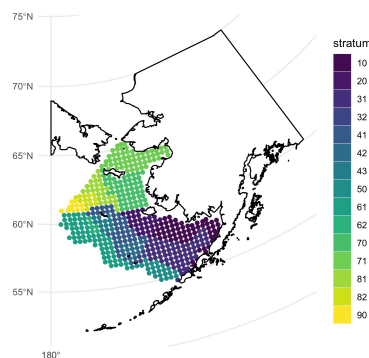

Figure 2: Survey replicated stations.

The code segment below will recreate the above figure 2.

```
# explore stations in the survey replicated data:
station_info

# first convert the station_info object into a shapefile for mapping:
station_sf        <- convert2shp(station_info)
station_sf$stratum <- factor(station_sf$stratum)

# plot the stations:
p <- plot_stations_basemap(sfIN = station_sf,fillIN = "subregion",colorIN = "subregion") +
  scale_color_viridis_d(begin = .4,end=.6) +
  scale_fill_viridis_d(begin = .4,end=.6)
p

p2 <- plot_stations_basemap(sfIN = station_sf,fillIN = "stratum",colorIN = "stratum") +
  scale_color_viridis_d() +
```

```r
    scale_fill_viridis_d()
  p2

  if(update.figs)  ggsave(file="Figs/stations.jpg",width=5,height=5)
```

Now let's explore the survey replicated data in more detail and use it to create a cold pool index for each simulation and hindcast scenario x model x CMIP combination.

```r
  # now create plots of average BT during four time periods
  time_seg   <- list('2010-2020' = c(2010:2020),
                     '2021-2040' = c(2021:2040),
                     '2041-2060' = c(2041:2060),
                     '2061-2080' = c(2061:2080),
                     '2081-2099' = c(2081:2099))


  # Collate mean values across timeperiods and simulations:
  m_set      <- c(18,19);aclim[m_set]
  mn_var_all <- get_mn_srvy_var(modset = aclim[m_set],varUSE="temp_bottom5m")
  mn_var_sf  <- convert2shp(mn_var_all%>%filter(!is.na(mnval)))

  p3         <- plot_stations_basemap(sfIN = mn_var_sf,
                            fillIN = "mnval",
                            colorIN = "mnval",
                            sizeIN=.3) +
    facet_grid(simulation~time_period)+
    scale_color_viridis_c()+
    scale_fill_viridis_c()+
    guides(
      color =  guide_legend(title="Bottom T (degC)"),
      fill  =  guide_legend(title="Bottom T (degC)")) +
    ggtitle(substr(aclim[m_set[1]],1,23))

  # This is slow but it works (repeat dev.new() twice if in Rstudio)...
  dev.new()
  p3
  if(update.figs)  ggsave(file="Figs/mn_BT.jpg",width=8,height=5)

  graphics.off()
```

2.3. Explore "weekly strata" `.nc` files in `R()`

The next set of indices to will explore are the weekly strata-specific values for each variable.These are stored in the `ACLIMregion_B10K-[version_CMIPx_GCM_RCP].nc` in each scenario folder.

```r
# list of the scenario x GCM downscaled ACLIM indices
for(k in aclim)
  cat(paste(k,"\n"))
info<-function(model_list=aclim[1:2],type=1){
  for(mod in model_list){
    if(type==1) ncfl    <- file.path(mod,paste0(reg_txt,mod,".nc"))
    if(type==2) ncfl    <- file.path(mod,paste0(reg_txt,mod,".nc"))
    if(type!%in%1:2) message("error")
     nc       <- nc_open(file.path(data_path,"roms_for_public",ncfl))
```
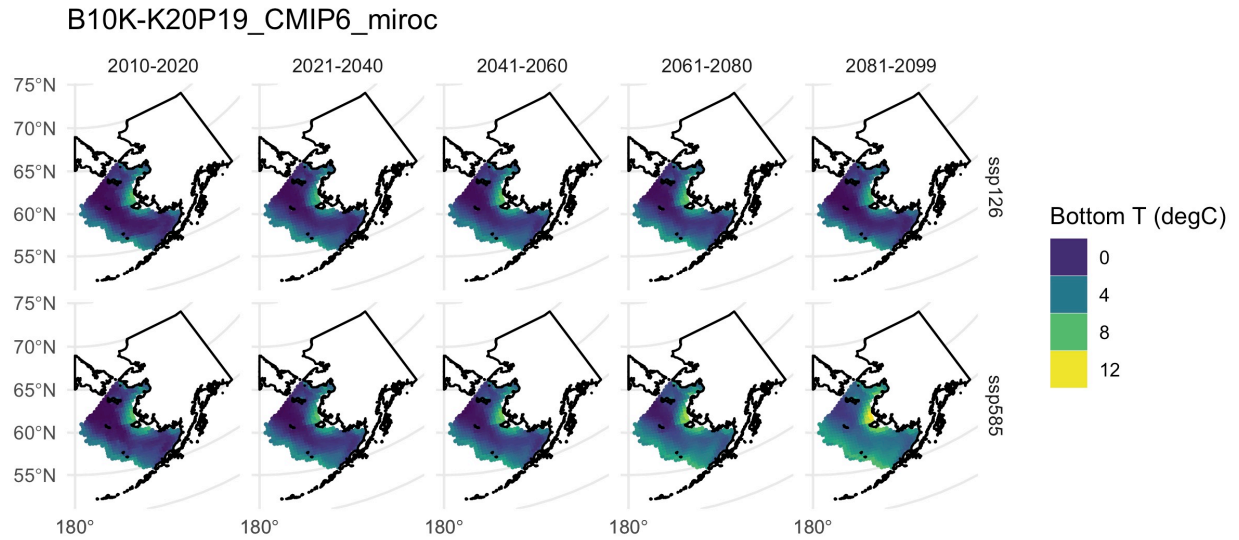
Figure 3: Bottom temperature projections under differing SSP126 (top row) and SSP585 (bottom row)

```r
      # get the timestep for each value:
      t        <- as.POSIXct(nc$var[[5]]$dim[[2]]$vals,
                    origin = substr(nc$var[[k]]$dim[[2]]$units,15,36),
                    tz = "GMT")
      Start = t[1],
      End   = rev(t)[1],
      nvars = length(vars)
       # get the area for each region:
      area   <- ncvar_get(nc, varid = "region_area")
      # get the full list of variables:
      vars   <- names(nc$var)
  }



}

# open a "region" or strata specific nc file:
i        <- 2
ncfl     <- file.path(aclim[i],paste0(reg_txt,aclim[i],".nc"))
nc       <- nc_open(file.path(data_path,"roms_for_public",ncfl))

# get the area for each region:
area   <- ncvar_get(nc, varid = "region_area")
# get the full list of variables:
vars   <- names(nc$var)
```

```r
# get the variable index number:
k       <- grep("temp_bottom5m", vars)

#region_names: 1 = EBS, 2 = SEBS, 3 = NEBS_US
nc$var[[k]]$dim[[1]]$name  # region
nc$var[[k]]$dim[[2]]$name  # ocean_time

# get the full list of survey strata names:
strata <- (nc$var[[k]]$dim[[1]]$vals)

# get the timestep for each value:
t       <- as.POSIXct(nc$var[[k]]$dim[[2]]$vals,
                 origin = substr(nc$var[[k]]$dim[[2]]$units,15,36),
                 tz = "GMT")
# get the values as a matrix of strata (22) x time (3908)
val     <- ncvar_get(nc, varid = vars[k])

NEBS_strata <- c(71,70,81,82,90)
SEBS_strata <- c(10,31,32,50,
                 20,42,61,62,43)

for(s in 1:length(strata)){
  tmp     <- data.frame(
    #scenario =
    strata          = strata[s],
    strata_area_km2 = area[s],
    time    = t,
    var     = vars[k],
    val     = val[s,],
    units   = nc$var[[k]]$units,
    long_name = nc$var[[k]]$longname)
  if(s == 1) val_long <- tmp
  if(s !=1 ) val_long <- rbind(val_long,tmp)
}

val_long$basin <- "Other"
val_long$basin[abs(val_long$strata)%in%NEBS_strata] <- "NEBS"
val_long$basin[abs(val_long$strata)%in%SEBS_strata] <- "SEBS"
val_long$basin  <- factor(val_long$basin, levels =c("NEBS","SEBS","Other"))
val_long$strata <-  factor(val_long$strata,
                           levels=(unique(strata)))

# now plot the data:

ggplot(data = val_long) +
  geom_line(aes(x=time,y=val,color= strata),alpha=.8)+
  facet_grid(basin~.)+
  ylab(val_long$units[1])+
  ggtitle(substr( dirlist[i],18,nchar( dirlist[i])-3))+
  theme_minimal()

# To get the average value for a set of strata, weight the val by the area:
```

```r
mn_NEBS <-getAVGnSUM(strataIN = NEBS_strata, dataIN=val_long)
mn_NEBS$basin = "NEBS"

mn_SEBS <-getAVGnSUM(strataIN = SEBS_strata, dataIN=val_long)
mn_SEBS$basin = "SEBS"


ggplot(data = rbind(mn_NEBS,mn_SEBS)) +
 geom_line(aes(x=time,y=mn_val,color=basin),alpha=.8)+
 geom_smooth(aes(x=time,y=mn_val,color=basin),formula = y ~ x, se = T)+
 facet_grid(basin~.)+
  scale_color_viridis_d(begin=.4,end=.8)+
 ylab(val_long$units[1])+
 ggtitle( dirlist[i])+
 theme_minimal()
```

3. Funding and acknowledgments (needs updating):

PLEASE Include a statement like the following one in your acknowledgements section:

For some of the integrated papers the following maybe should also be added:

4. Helpful links and further reading:

4.1 Citations for GCMs and carbon scenarios:

CMIP3 (BSIERP global climate model runs):

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. Bull. Amer. Meteor. Soc., 88, 1383–1394.

CMIP5 (ACLIM global climate model runs):

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012:Anoverview of CMIP5 and the experiment design. Bull. Amer. Meteor. Soc., 93, 485–498.

CMIP6 and SSPs (ACLIM2 global climate model runs):

ONeill, B. C., C. Tebaldi, D. P. van Vuuren, V. Eyring, P. Friedlingstein, G. Hurtt, R. Knutti, E. Kriegler, J.-F. Lamarque, J. Lowe, G. A. Meehl, R. Moss, K. Riahi, and B. M. Sanderson. 2016. The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. Geoscientific Model Development 9:3461–3482.

4.2 Weblinks for further reading:

- Explore annual indices of downscaled projections for the EBS: **ACLIM indices**
- To view climate change projections from CMIP5 (eventually CMIP6):**ESRL climate change portal**

4.3 Additional information on Hindcast and Projection Models (needs updating)

CORE-CFSR (1976-2012)

This is the hindcast for the Bering Sea and is a combination of the reconstructed climatology from the **CLIVAR** Co-ordinated Ocean-Ice Reference Experiments (CORE) Climate Model (1969-2006) the **NCEP** Climate Forecast System Reanalysis is a set of re-forecasts carried out by NOAA's National Center for Environmental Prediction (NCEP). See **CFS-R** for more info.

CCCMA(2006-2039; AR4 SRES A1B)

Developed by the Canadian Centre for Climate Modelling and Analysis, this is also known as the CGCM3/T47 model. This model showed the greatest warming over time compared to other models tested by PMEL. See more data the **AOOS:CCCMA portal**.

ECHOG(2006-2039; AR4 SRES A1B)

The ECHO-G model from the Max Planck Institute in Germany This model showed the least warming over time compared to other models tested by PMEL. See more data the AOOS:ECHO-G portal.

GFDL (2006-2100; AR5 RCP 4.5, 8.5, SSP126,SSP585)

The NOAA Geophysical Fluid Dynamics Laboratory **GFDL** has lead development of the first Earth System Models (ESMs), which like physical climate models, are based on an atmospheric circulation model coupled with an oceanic circulation model, with representations of land, sea ice and iceberg dynamics; ESMs additionally incorporate interactive biogeochemistry, including the carbon cycle. The ESM2M model used in this project is an evolution of the prototype EMS2.1 model, where pressure-based vertical coordinates are used along the developmental path of GFDL's Modular Ocean Model version 4.1 and where the land model is more adavanced (LM3) than in the previous ESM2.1

MIROC(2006-2039; AR4 SRES A1B; 2006-2100 RCP4.5, RCP8.5, SSP585, SSP126)

The Model for Interdisciplinary Research on Climate (MIROC)-M model developed by a consortium of agencies in Japan []. Compared to other models tested by PMEL, MIROC-M was intermediate in degree of warming over the Bering Sea shelf for the first half of the 21st century. See more data the AOOS:MIROC portal.

ACLIM_data & ACLIM_scripts

This folder contains the most recent queries from the BEAST based on the scripts in ACLIM_scripts. The queries create annual indices (e.g., mean summer bottom temperature for the EBS survey area) which are station specific(Station_modelname) and based on the mean sampling date for each station, or based on weekly values (under folder weekly; are "clipped to the survey area"). These indices were cleaned and aggregated in the .Rdata file `ROMSNPZ_indices.Rdata` using the script `createROMSNPZ.R.`