Getting Started with Bering10K ROMSNPZ Level 2 & 3 indices

Contents

Figure 1: .

1. ACLIM code and online repo and ROMSNPZ data 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 Bering10K ROMSNPZ documentation can be accessed on the main **documentation** webpage. The webpage is maintained by Kelly Kearney and regularly updated with new documentation, including the following core documents (ACLIM members can also find these linked in the **00_ACLIM_shared/02_Data** folder):

- **The Bering10K Dataset documentation**: A pdf describing the dataset, including:

  1. A description of the various simulations (base model versions, parent model forcing datasets, and biological modules) and the output naming scheme for each.

  2. A tutorial on working with the curvilinear, sigma-coordinate ROMS spatial grid, aimed at those who will be working with data on the native grid.

  3. An overview of the ACLIM index dataset; this is a set of time series derived from the Bering10K output and intended for Alaska Climate Integrated Modeling (ACLIM) collaborators.

- **Bering10K Simulaton Variables**: A spreadsheet listing all simulations and the archived output variables associated with each, updated periodically as new simulations are run or new variables are made available. Note that this includes both data available on both public and private servers (see below). Please also see the Literature page for a list scientific publications related to the model, including model description and skill assessment.

1.1. Downscaled models and carbon scenarios

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

Table 1: Summary of ROMSNPZ downscaled model runs

CMIP	GCM	Scenario	Def	Years	Model	Source	Status
	CORECFS	Reanalysis	Hindcast	1970 - 2018	H16	IEA/ACLIM	Public
	CORECFS	Reanalysis	Hindcast	1970 - 2020	K20	MAPP/IEA/ACLIM	Public
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
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

CMIP	GCM	Scenario	Def	Years	Model	Source	Status
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. Guildlines for use and citation of the data

It is strongly recommended that you include at least one (ideally multiple) authors from the ROMSNPZ team (Drs. Hermann, Cheng, Kearney, Pilcher) 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.

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 scale general circulation models (GCMs) under two CMIP (Coupled Model Intercomparison Project]) phase 5 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 2018 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 the most recent CMIP phase 6.

- **Hindcast (1979-2020 hindcast with OA dynamics used in ACLIM 2.0):**

  Kearney, K., A. Hermann, W. Cheng, I. Ortiz, and K. Aydin. 2020. A coupled pelagic-benthic-sympagic biogeochemical model for the Bering Sea: documentation and validation of the BESTNPZ model (v2019.08.23) within a high-resolution regional ocean model. Geoscientific Model Development 13:597–650.

  Pilcher, D. J., D. M. Naiman, J. N. Cross, A. J. Hermann, S. A. Siedlecki, G. A. Gibson, and J. T. Mathis. 2019. Modeled Effect of Coastal Biogeochemical Processes, Climate Variability, and Ocean Acidification on Aragonite Saturation State in the Bering Sea. Frontiers in Marine Science 5:1–18.

- **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)

2. Get ACLIM code (Step 1)

IMPORTANT
The ACLIM indices and ROMSNPZ simulations are stored as netcdf files (.nc) format in the Data folder of the ACLIM shared google drive (section 2.3) or available on the new PMEL web-based portal (see section 2.2 below). 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.

The code for accessing plotting the data is available in the ACLIM2 github repository: **ACLIM ROM-SNPZ Repo: github.com/kholsman/ACLIM2**. This code will open the netcdf (.nc) files in R and save them on your local drive as .Rdata files. The code also will 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**.

There are two ways to put the repo on your local drive 1) use R to download it and 2) manually download a zipped version of the repo. **Once you've unzipped the folder, if you have Rstudio installed you can double click on the `ACLIM2/ACLIM2.Rproj` and use Rstudio to manage your plotting and files (recommended).**

2.1 Option 1: Use R to download from ACLIM2 github repo:

```r
# Specify the download directory
main_nm       <- "ACLIM2"
download_path <- path.expand("~/desktop")
dest_fldr     <- file.path(download_path,main_nm)

url           <- "https://github.com/kholsman/ACLIM2/archive/main.zip"
dest_file     <- file.path(download_path,paste0(main_nm,".zip"))
download.file(url=url, destfile=dest_file)

# unzip the .zip file
setwd(download_path)
unzip (dest_file, exdir = "./",overwrite = T)

#rename the unzipped folder from ACLIM2-main to ACLIM2
file.rename(paste0(main_nm,"-main"), main_nm)
setwd(main_nm)
```
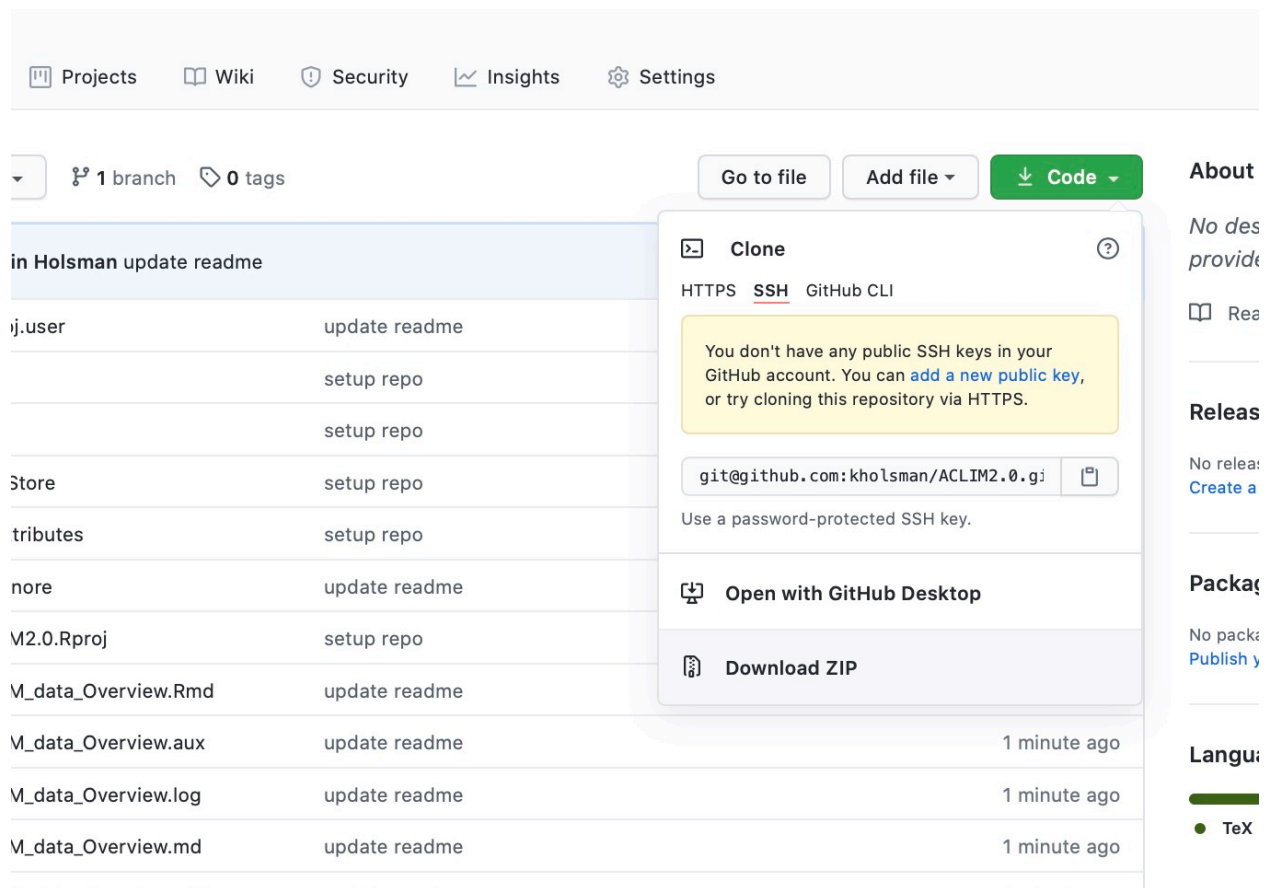
2.2 Option 2: Manually download from ACLIM2 github repo

Select `Download ZIP` from the upper right hand side of the repo page :**github.com/kholsman/ACLIM2** and save it to your local directory: `~[YOURPATH]/ACLIM2`.

3. Get ROMSNPZ data (Step 2)

3.1 Available data sources

There are presently two sources of ROMSNPZ level 2 and level 3 post-processed datasets:

1) Public web-based ACLIM CMIP5 datasets

- `Level1` : (Empty; data not copied from Mox)
- `Level2` : (full grid, rotated to lat lon from the native ROMSNPZ grid, weekly averages)
  - `Bottom 5m` : subset of variables from the bottom 5 m of the water column
  - `Bottom 5m` : subset of variables for the surface 5 m of the water column
  - `Integrated`: watercolumn integrated averages or totals for various variables
  - `Level3`: two post-processed datasets
    * `ACLIMsurveyrep-x.nc.`: Survey replicated (variables "sampled" at the average location and date that each groundfish survey is sampled)*(Note that the resampling stations need to be removed before creating bottom temperature maps)*

  - `ACLIMregion-xnc.`:weekly variables averaged for each survey strata *(Note that area (km2) weighting should be used to combine values across multiple strata)*

2) ACLIM google drive CMIP6 datasets (embargoed; *ACLIM members Only*)

- `Level1`: (Empty; data not copied from Mox)
- `Level2`: (Empty; data not copied from Mox)
- `Level3`: two post-processed datasets
  - `ACLIMsurveyrep-x.nc.`: Survey replicated (variables "sampled" at the average location and date that each groundfish survey is sampled)*(Note that the resampling stations need to be removed before creating bottom temperature maps)*

  - `ACLIMregion-xnc.`:weekly variables averaged for each survey strata *(Note that area (km2) weighting should be used to combine values across multiple strata)*

For all files the general 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.

3.1.1 Option 1: Public web-based ACLIM data (hindcasts & CMIP5 projections)

The public web-based ACLIM data (hindcasts & CMIP5 projections) option is available for Level3 and Level2 CMIP5 public data, it is not yet available for the embargoed CMIP6 data but through ACLIM2.0 will eventually be used to host that as well.

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 exploration, query, mapping, and plotting tool

- ERDAPP ACLIM data access tool

- THREDDS ACLIM direct access to Level 2 and 3 netcdf files

3.1.2 Option 2: (*ACLIM members Only*) Access CMIP6 (embargoed) L3 data

Public CMIP5 and embargoed CMIP6 Level 3 netcdf (.nc) files are saved in the shared ACLIM data folder (note: Level 2 files are too large for the google drive but are available by request from Kelly Kearney.

IMPORTANT Please note that while the CMIP5 set is now public (Hermann et al. 2019; section 2.2) **the CMIP6 suite is under embargo for QAQC and should not be shared outside of the ACLIM group**. The ROMSNPZ team (Drs. Hermann, Cheng, Kearney, Pilcher, Adyin) 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. The ROMSNPZ team has made these runs available to ACLIM2 members in order to accelerate coupling to biological and social and economic models, thus out of professional courtesy please do not publish the data without permission from **all** ROMSNPZ team members, it is strongly advised that some or multiple ROMSNPZ team members be included as co-authors to ensure proper application and use of the ROMSNPZ data.

For most applications you can use the ACLIM level3 post-processed indices available on the shared ACLIM drive in the root google drive data folder: **00_ACLIM_shared>02_DATA**.

The `Newest` 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 (km2) weighting should be used to combine values across multiple strata)*

There are two folders that need to be copied into the ACLIM2 folder on your computer under '`~[YOURPATH]/ACLIM2/Data/in/`:

1) **00_ACLIM_shared>02_DATA>Newest**. This folder contains a folder called `roms_for_aclim` with all the ACLIM Level3 indices for model simulations available to ACLIM members.

2) **00_ACLIM_shared>02_DATA>Map_layers.zip**. This file needs to be unzipped after you download it to your local folder. It contains (large) base maps for the code below including `shp_files` and `geo_tif` folders.
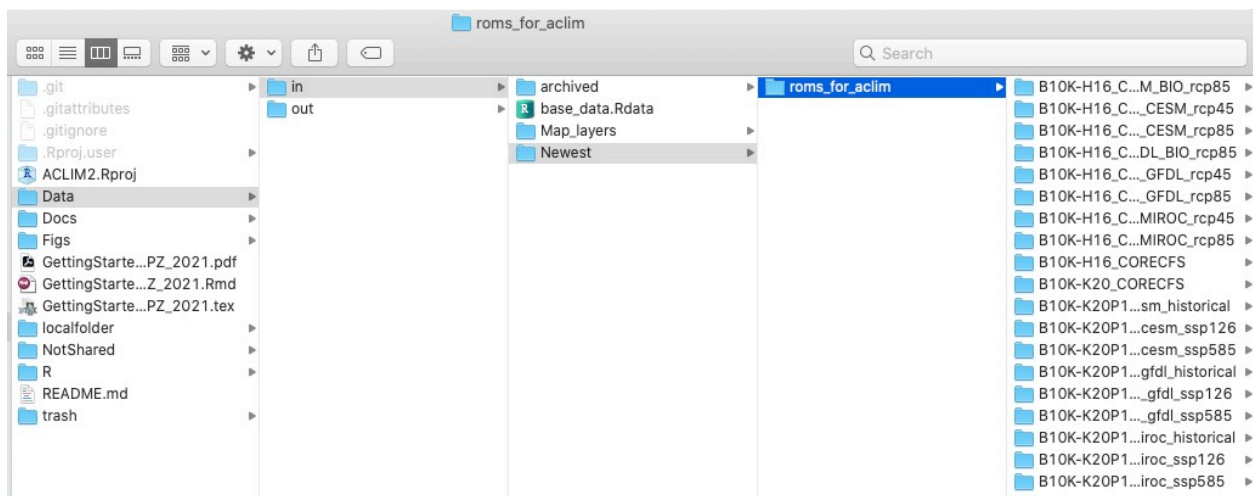

Figure 2: Your local `ACLIM2/Data` directory should look something like this when you are done downloading the data and unzipping it.

3.2 Access and save the data

The below code will extract variables from the Level 2 and Level 3 netcdf files (`.nc`) and save them as compressed `.Rdata` files on your local Data/in/Newest/Rdata folder.

3.2.1 Setup up the R worksace

First let's get the workspace set up, will we step through an example downloading the hindcast and a single projection (CMIP5 MIROC rcp8.5) but you can loop the code below to download the full set of CMIP5 projections.

```
# --------------------------------------
# SETUP WORKSPACE
tmstp  <- format(Sys.time(), "%Y_%m_%d")
```

7

```
    main    <- getwd()  #"~/GitHub_new/ACLIM2
    source("R/make.R")
    # ------------------------------------
```

Let's take a look at the available online datasets:

```
    # preview the datasets on the server:
    url_list <- tds_list_datasets(thredds_url = ACLIM_data_url)

    #display the full set of datasets:
    cat(paste(url_list$dataset,"\n"))
```

```
## Constants/
##  B10K-H16_CMIP5_CESM_BIO_rcp85/
##  B10K-H16_CMIP5_CESM_rcp45/
##  B10K-H16_CMIP5_CESM_rcp85/
##  B10K-H16_CMIP5_GFDL_BIO_rcp85/
##  B10K-H16_CMIP5_GFDL_rcp45/
##  B10K-H16_CMIP5_GFDL_rcp85/
##  B10K-H16_CMIP5_MIROC_rcp45/
##  B10K-H16_CMIP5_MIROC_rcp85/
##  B10K-H16_CORECFS/
##  B10K-K20_CORECFS/
##  files/
```

3.2.2 Download Level 2 data

First we will explore the Level 2 bottom temperature data on the ACLIM Thredds server using the H16 hindcast and the H16 (CMIP5) projection for MIROC under rcp8.5. The first step is to get the data urls:

```
    # define the simulation to download:
    cmip <- "CMIP5"      # Coupled Model Intercomparison Phase
    GCM  <- "MIROC"      # Global Circulation Model
    rcp  <- "rcp85"      # future carbon scenario
    mod  <- "B10K-H16"   # ROMSNPZ model
    hind <- "CORECFS"    # Hindcast

    # define the projection simulation:
    proj  <- paste0(mod,"_",cmip,"_",GCM,"_",rcp)
    hind  <- paste0(mod,"_",hind)

    # get the url for the projection and hindcast datasets:
    proj_url        <- url_list[url_list$dataset == paste0(proj,"/"),]$path
    hind_url        <- url_list[url_list$dataset == paste0(hind,"/"),]$path

    # preview the projection and hindcast data and data catalogs (Level 1, 2, and 3):
    proj_datasets  <- tds_list_datasets(thredds_url = proj_url)
    hind_datasets  <- tds_list_datasets(thredds_url = hind_url)

    # get url for the projection and hindcast Level 2 and Level 3 catalogs
    proj_l2_cat    <- proj_datasets[proj_datasets$dataset == "Level 2/",]$path
    proj_l3_cat    <- proj_datasets[proj_datasets$dataset == "Level 3/",]$path
    hind_l2_cat    <- hind_datasets[hind_datasets$dataset == "Level 2/",]$path
    hind_l3_cat    <- hind_datasets[hind_datasets$dataset == "Level 3/",]$path
```

8

```
    hind_l2_cat
```

[1] "https://data.pmel.noaa.gov/aclim/thredds/B10K-H16_CORECFS/Level2.html"

Now that we have the URLs let's take a look at the available Level2 datasets (currently temperature only, other variables available by request to Kelly Kearney:

- `Bottom 5m` : bottom water temperature at 5 meters
- `Surface 5m` : surface water temperature in the first 5 meters
- `Integrated` : Integrated water column averages for various NPZ variables

```r
# preview the projection and hindcast Level 2 datasets:
proj_l2_datasets  <- tds_list_datasets(proj_l2_cat)
hind_l2_datasets  <- tds_list_datasets(hind_l2_cat)
proj_l2_datasets$dataset
```

[1] "Bottom 5m" "Surface 5m" "Integrated"

```r
# get url for bottom temperature:
proj_l2_BT_url    <- proj_l2_datasets[proj_l2_datasets$dataset == "Bottom 5m",]$path
hind_l2_BT_url    <- hind_l2_datasets[hind_l2_datasets$dataset == "Bottom 5m",]$path
proj_l2_BT_url
```

[1] "https://data.pmel.noaa.gov/aclim/thredds/B10K-H16_CMIP5_MIROC_rcp85/Level2.html?dataset=B10K-H1(

We can't preview the Level 3 datasets in the same way but they are identical to those in the google drive and include two datasets

- `ACLIMsurveyrep_B10K-H16_CMIP5_CESM_BIO_rcp85.nc` : NMFS Groundfish summer NBS and EBS survey replicated values for 60+ variables
- `ACLIMregion_B10K-H16_CMIP5_CESM_BIO_rcp85.nc` : weekly strata averages for 60+ variables

```r
weekly_vars  # list of possible variables in the ACLIMregion_ files
```

```
##  [1] "region_area"        "Ben"                "DetBen"
##  [4] "Hsbl"               "IceNH4"             "IceNO3"
##  [7] "IcePhL"             "aice"               "hice"
## [10] "shflux"             "ssflux"             "Cop_integrated"
## [13] "Cop_surface5m"      "EupO_integrated"    "EupO_surface5m"
## [16] "EupS_integrated"    "EupS_surface5m"     "Iron_bottom5m"
## [19] "Iron_integrated"    "Iron_surface5m"     "Jel_integrated"
## [22] "Jel_surface5m"      "MZL_integrated"     "MZL_surface5m"
## [25] "NCaO_integrated"    "NCaO_surface5m"     "NCaS_integrated"
## [28] "NCaS_surface5m"     "NH4_bottom5m"       "NH4_integrated"
## [31] "NH4_surface5m"      "NO3_bottom5m"       "NO3_integrated"
## [34] "NO3_surface5m"      "PhL_integrated"     "PhL_surface5m"
## [37] "PhS_integrated"     "PhS_surface5m"      "prod_Cop_integrated"
## [40] "prod_EupO_integrated" "prod_EupS_integrated" "prod_Eup_integrated"
## [43] "prod_Jel_integrated" "prod_MZL_integrated" "prod_NCaO_integrated"
## [46] "prod_NCaS_integrated" "prod_NCa_integrated" "prod_PhL_integrated"
## [49] "prod_PhS_integrated" "salt_surface5m"     "temp_bottom5m"
## [52] "temp_integrated"    "temp_surface5m"     "uEast_bottom5m"
## [55] "uEast_surface5m"    "vNorth_bottom5m"    "vNorth_surface5m"
## [58] "fracbelow0"         "fracbelow1"         "fracbelow2"
```

Now we can download a subset of the Level2 data (full 10KM Lat Lon re-gridded data), here with an example of sampling on Aug 1 of each year:

```r
# Currently available Level 2 variables
 dl      <- proj_l2_datasets$dataset  # datasets

# variable list
svl <- list(
       'Bottom 5m' = "temp",
       'Surface 5m' = "temp",
       'Integrated' = c("EupS","Cop","NCaS") )


 # preview the variables, timesteps, and lat lon in each dataset:
 l2_info <- scan_l2(ds_list = dl,sim_list = "B10K-H16_CORECFS" )

 names(l2_info)
 l2_info[["Bottom 5m"]]$vars
 l2_info[["Surface 5m"]]$vars
 l2_info[["Integrated"]]$vars
 max(l2_info[["Integrated"]]$time_steps)
 l2_info[["Integrated"]]$years


 # Simulation list:
 # --> --> Tinker:add additional projection scenarios here
 sl <- c(hind, proj)

 # Currently available Level 2 variables
 dl      <- proj_l2_datasets$dataset  # datasets

 # variables to pull from each data set
 # --> --> Tinker: try subbing in other Integrated variables
 # (l2_info[["Integrated"]]$vars) into the third list vector
 svl <- list(
   'Bottom 5m' = "temp",
   'Surface 5m' = "temp",
   'Integrated' = c("EupS","Cop","NCaS") )


 # Let's sample the model years as close to Aug 1 as the model timesteps run:
 # --> --> Tinker - try a different date
 tr          <- c("-08-1 12:00:00 GMT")

 # grab nc files from the aclim server and convert to rdatafiles with the ID Aug1
 get_l2(
   ID          = "_Aug1",
   ds_list     = dl,
   trIN        = tr,
   sub_varlist = svl,
   sim_list    = sl  )
```

3.2.3 Download Level 3 data

Now let's grab some of the Level 3 data and store it in the Data/in/Newest/Rdata folder. This is comparatively faster because Level 3 files are already post-processed to be in the ACLIM indices format and are

relatively small:

```
# Simulation list:
# --> --> Tinker:add additional projection scenarios here
sl <- c(hind, proj)

# variable list
# --> --> Tinker:add additional variables to varlist
vl <- c(
        "temp_bottom5m",     # bottom temperature,
        "NCaS_integrated",   # Large Cop
        "Cop_integrated",    # Small Cop
        "EupS_integrated")   # Shelf  euphausiids

# convert  nc files into a long data.frame for each variable
# three options are:
# ------------------------------------

# opt 1: access nc files remotely (fast, less local storage needed)
get_l3(web_nc = TRUE, download_nc = F,
       varlist = vl,sim_list = sl)

# opt 2:  download nc files then access locallly:
get_l3(web_nc = TRUE, download_nc = T,
       local_path = file.path(local_fl,"aclim_thredds"),
       varlist = vl,sim_list = sl)

 # opt 3:  access existing nc files locally:
get_l3(web_nc = F, download_nc = F,
       local_path = file.path(local_fl,"aclim_thredds"),
       varlist = vl,sim_list = sl)
```

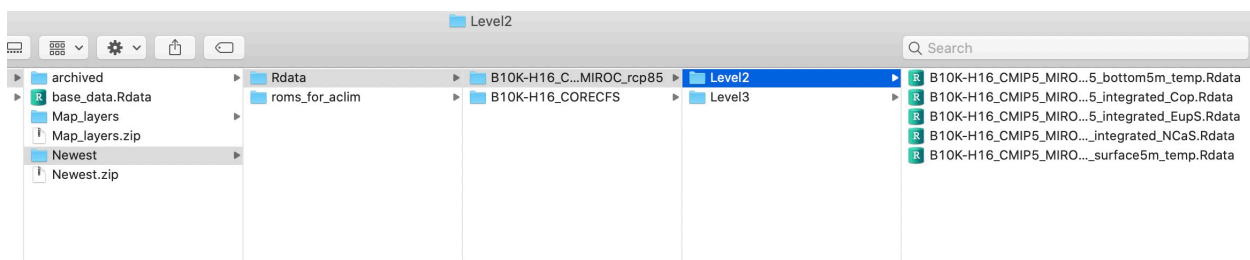### 3.2.4 ACLIM only: Convert CMIP6 Level 3 .nc −> .Rdata

Figure 3: The final folder structure on your local drive in `Data/in/Newest` should look something like this.

Now let's convert the CMIP6 Level3 .nc files to .Rdata files (as in section 3.2.3)

```
# preview the available CMIP6 data
dir(file.path(local_fl,"roms_for_aclim"))

# variable list
vl          <- c(
            "temp_bottom5m",
```

11

```
                 "NCaS_integrated", # Large Cop
                 "Cop_integrated",  # Small Cop
                 "EupS_integrated") # Euphausiids

   # define the simulation to download:
     cmip <- "CMIP6"      # Coupled Model Intercomparison Phase
     GCM  <- "MIROC"      # Global Circulation Model
     rcp  <- "ssp585"      # future carbon scenario
     mod  <- "B10K-K20P19" # ROMSNPZ model
     hind <- "CORECFS"       # Hindcast

     # define the projection simulation:
     proj  <- paste0(mod,"_",cmip,"_",GCM,"_",rcp)
     hind  <- paste0("B10K-K20","_",hind)
     sl    <- c(hind, proj)

     # opt 3:  convert subset of nc files to rdata files for analysis:
     get_l3(web_nc = F, download_nc = F,
          local_path = file.path(local_fl,"roms_for_aclim"),
          varlist = vl,sim_list = sl)
```

4. Explore indices & plot the data

4.1 Level 3 indices:

Level 3 indices can be used to generate seasonal, monthly, and annual indices (like those reported in Reum et al. 2020), Holsman et al. 2020). In the section 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. In section 3 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. The following examples show how to analyze and plot the ACLIM indices from the .Rdata files created in the previous step 3.

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 **Al Hermann** or **Kelly Kearney**. 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**. See Section 1 above for more detail.

4.1.1 Explore Level 3 data catalog

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)*

First run the below set of code to set up the workspace:

```r
# -------------------------------------
# SETUP WORKSPACE
tmstp  <- format(Sys.time(), "%Y_%m_%d")
main   <- getwd()  #"~/GitHub_new/ACLIM2
source("R/make.R")
# -------------------------------------

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

embargoed # not yet public or published
public    # published runs (CMIP5)

# 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

# explore stations in the survey replicated data:
head(station_info)
```

4.1.2 Level 3: Spatial indices (survey replicated)

Let's start b exploring the survey replicated values for each variable. Steps 2 and 3 generated the Rdata files that are stored in the `ACLIMsurveyrep_B10K-[version_CMIPx_GCM_RCP].Rdata` in each corresponding simulation folder.
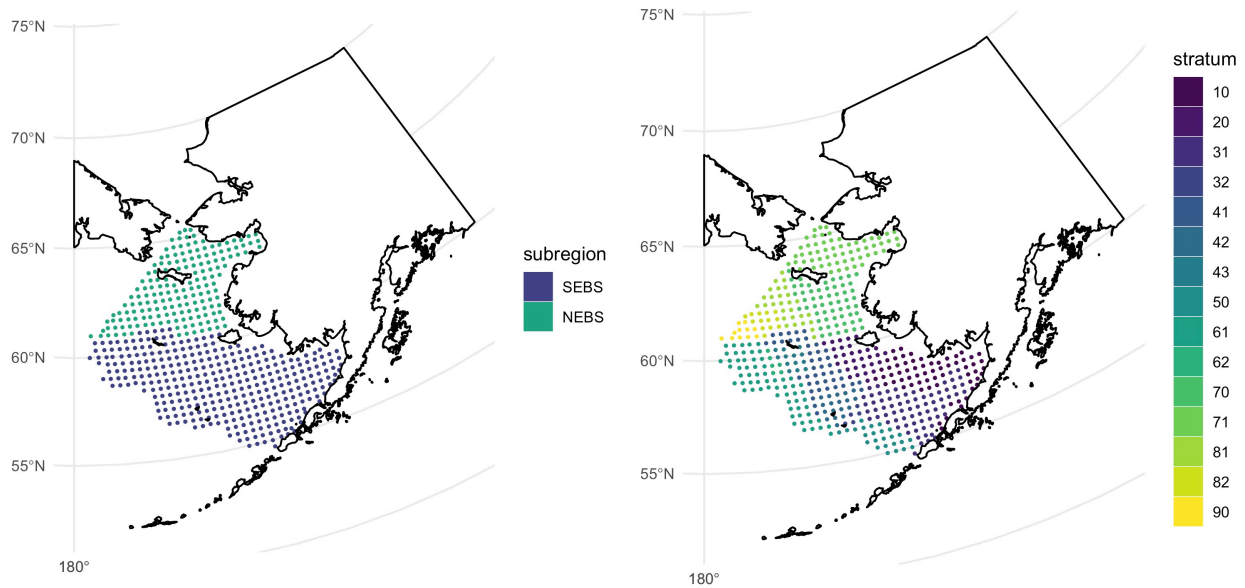
The code segment below will recreate the above figures.

```r
# 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 = .2,end=.6) +
  scale_fill_viridis_d(begin  = .2,end=.6)

if(update.figs){
  p
  ggsave(file=file.path(main,"Figs/stations_NS.jpg"),width=5,height=5)
 }

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

if(update.figs){
  p2
ggsave(file=file.path(main,"Figs/stations.jpg"),width=5,height=5)}
```

Now let's explore the survey replicated data in more detail and use to plot bottom temperature.

```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))
```

14

```r
# View an individual variable (e.g., Bottom Temp)
# -------------------------------------------------------
head(srvy_vars)
head(aclim)

# assign the simulation to download
# --> --> Tinker: try selecting a different set of models to compare
sim         <-"B10K-H16_CMIP5_MIROC_rcp85"

# open a "region" or strata specific nc file
fl          <- file.path(sim,paste0(srvy_txt,sim,".Rdata"))

# load object 'ACLIMsurveyrep'
load(file.path(main,Rdata_path,fl))

# Collate mean values across timeperiods and simulations
# -------------------------------------------------------
m_set       <- c(9,7,8)
ms          <- aclim[m_set]

# create local rdata files (opt 1)
if(!file.exists(fl))
  get_l3(web_nc = TRUE, download_nc = F,
       varlist = vl,sim_list =sim )

# get the mean values for the time blocks from the rdata versions
# will throw "implicit NA" errors that can be ignored
mn_var_all <- get_mn_rd(modset = ms ,varUSE="temp_bottom5m")

# convert results to a shapefile
mn_var_sf  <- convert2shp(mn_var_all%>%filter(!is.na(mnval)))
lab_t       <- ms[2]%>%stringr::str_remove("([^-])")

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(lab_t)

# This is slow but it works (repeat dev.new() twice if in Rstudio)...
dev.new()
p3

if(update.figs)  ggsave(file=file.path(main,"Figs/mn_BT.jpg"),width=8,height=6)

# graphics.off()
```
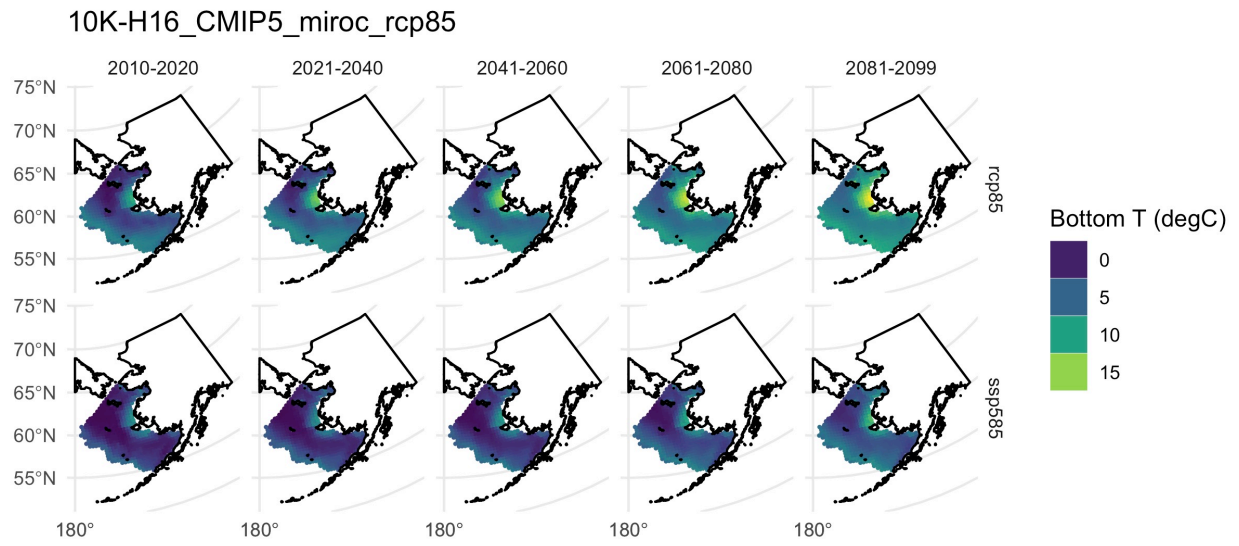
Figure 4: Bottom temperature projections under differing SSP126 (top row) and SSP585 (bottom row)

4.1.3 Level 3: Weekly strata averages

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
# View an individual variable (e.g., Bottom Temp)
# -----------------------------------------------------
weekly_vars
aclim
sim         <-"B10K-H16_CMIP5_MIROC_rcp85"

# open a "region" or strata specific nc file
fl          <- file.path(sim,paste0(reg_txt,sim,".Rdata"))

var_use    <- "temp_bottom5m"
vl         <- c(
                "temp_bottom5m",
                "NCaS_integrated", # Large Cop
                "Cop_integrated",  # Small Cop
                "EupS_integrated") # Euphausiids

# create local rdata files (opt 1)
if(!file.exists(fl))
  get_l3(web_nc = TRUE, download_nc = F,
        varlist = vl,sim_list = sim)

# load object 'ACLIMregion'
load(file.path(main,Rdata_path,fl))
tmp_var     <- ACLIMregion%>%filter(var == var_use)

# now plot the data:

p4 <- ggplot(data = tmp_var) +
  geom_line(aes(x=time,y=val,color= strata),alpha=.8)+
  facet_grid(basin~.)+
  ylab(tmp_var$units[1])+
  ggtitle( paste(sim,tmp_var$var[1]))+
  theme_minimal()
p4
 if(update.figs)  ggsave(file=file.path(main,"Figs/weekly_bystrata.jpg"),width=8,height=5)


# To get the average value for a set of strata, weight the val by the area:
mn_NEBS <- getAVGnSUM(strataIN = NEBS_strata, dataIN = tmp_var)
mn_NEBS$basin = "NEBS"
mn_SEBS <-getAVGnSUM(strataIN = SEBS_strata, dataIN = tmp_var)
mn_SEBS$basin = "SEBS"

p5 <- 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(tmp_var$units[1])+
        ggtitle( paste(sim,mn_NEBS$var[1]))+

        theme_minimal()
    p5
    if(update.figs)  ggsave(file=file.path(main,"Figs/weekly_byreg.jpg"),width=8,height=5)
```
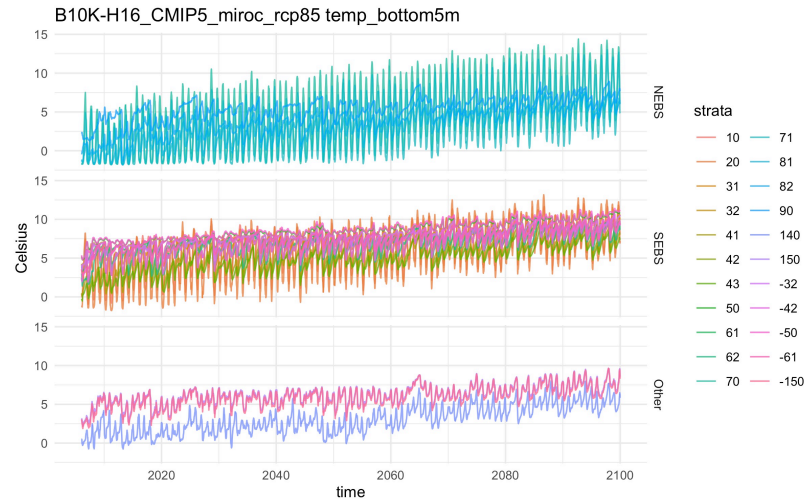

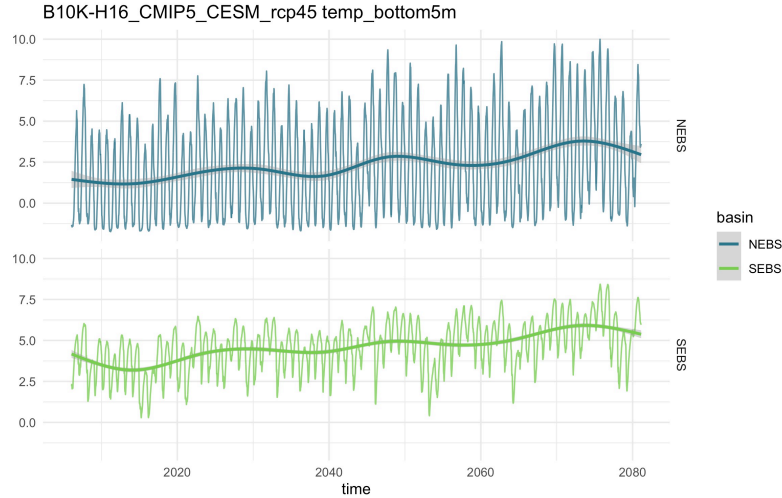Figure 5: Weekly indcices by sub-region

Figure 6: Weekly indcices by sub-region

4.1.4 Level 3: Seasonal averages

Now using a similar approach get the monthly mean values for a variable:

```
    sim <-"B10K-H16_CMIP5_MIROC_rcp85"

    # Set up seasons (this follows Holsman et al. 2020)
```

```r
    seasons <- data.frame(mo = 1:12,
                 season =factor("",
                   levels=c("Winter","Spring","Summer","Fall")))
    seasons$season[1:3]   <- "Winter"
    seasons$season[4:6]   <- "Spring"
    seasons$season[7:9]   <- "Summer"
    seasons$season[10:12] <- "Fall"


vl <- c(
              "temp_bottom5m",
              "NCaS_integrated", # Large Cop
              "Cop_integrated",  # Small Cop
              "EupS_integrated") # Euphausiids

# create local rdata files (opt 1)
if(!file.exists(fl))
  get_l3(web_nc = TRUE, download_nc = F,
      varlist = vl,sim_list = sim)

# open a "region" or strata specific  file
fl      <- file.path(sim,paste0(reg_txt,sim,".Rdata"))
load(file.path(main,Rdata_path,fl))

# get large zooplankton as the sum of euph and NCaS
tmp_var     <- ACLIMregion%>%
  filter(var%in%vl[c(2,3)])%>%
  group_by(time,strata,strata_area_km2,basin)%>%
  group_by(time,
        strata,
        strata_area_km2,
        basin,
        units)%>%
  summarise(val =sum(val))%>%
  mutate(var       = "Zoop_integrated",
        long_name ="Total On-shelf
        large zooplankton concentration,
        integrated over depth (NCa, Eup)")

rm(ACLIMregion)
head(tmp_var)

tmp_var$yr     <- strptime(as.Date(tmp_var$time),
                        format="%Y-%m-%d")$year + 1900
tmp_var$mo     <- strptime(as.Date(tmp_var$time),
                        format="%Y-%m-%d")$mon  + 1
tmp_var$jday   <- strptime(as.Date(tmp_var$time),
                        format="%Y-%m-%d")$yday + 1
tmp_var$season <- seasons[tmp_var$mo,2]

# To get the average value for a set of strata, weight the val by the area: (slow...)
mn_NEBS_season <- getAVGnSUM(
  strataIN = NEBS_strata,
```

```r
      dataIN = tmp_var,
      tblock=c("yr","season"))
  mn_NEBS_season$basin = "NEBS"
  mn_SEBS_season <- getAVGnSUM(
      strataIN = SEBS_strata,
      dataIN = tmp_var,
      tblock=c("yr","season"))
  mn_SEBS_season$basin = "SEBS"

  plot_data       <- rbind(mn_NEBS_season,mn_SEBS_season)

  # plot Fall values:
  p6 <- ggplot(data = plot_data%>%filter(season=="Fall") ) +
      geom_line(    aes(x = yr,y = mn_val,color=basin),alpha=.8)+
      geom_smooth( aes(x = yr,y = mn_val,color=basin),
                   formula = y ~ x, se = T)+
      facet_grid(basin~.)+
      scale_color_viridis_d(begin=.4,end=.8)+
      ylab(tmp_var$units[1])+
      ggtitle( paste(sim,"Fall",mn_NEBS_season$var[1]))+
      theme_minimal()
p6


if(update.figs)
  ggsave(file=file.path(main,"Figs/Fall_large_Zoop.jpg"),width=8,height=5)
```
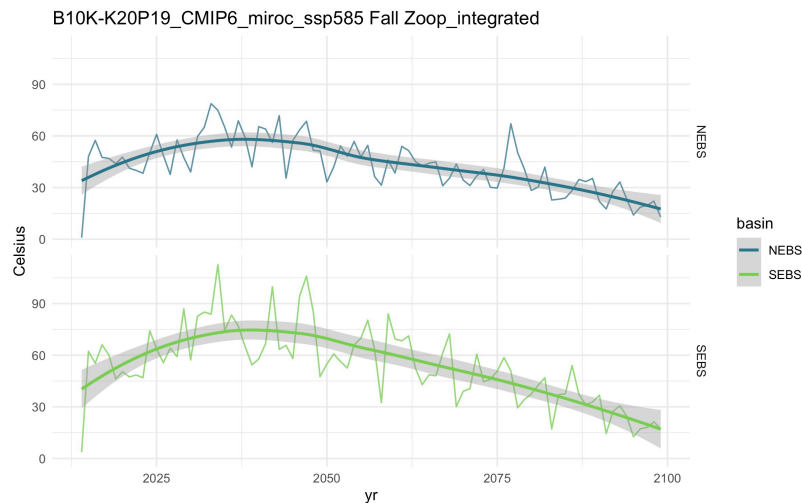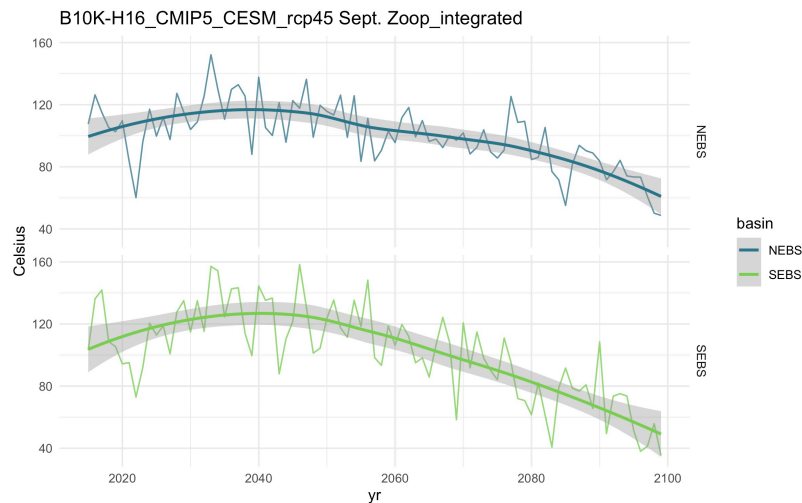

Figure 7: Large fall zooplankton integrated concentration

4.1.5 Level 3: Monthly averages

Using the same approach we can get monthly averages for a given variable:

```r
  # To get the average value for a set of strata, weight the val by the area: (slow...)
  mn_NEBS_season <- getAVGnSUM(
      strataIN = NEBS_strata,
```

```
      dataIN   = tmp_var,
      tblock   = c("yr","mo"))
  mn_NEBS_season$basin = "NEBS"
  mn_SEBS_season <- getAVGnSUM(
    strataIN = SEBS_strata,
    dataIN = tmp_var,
    tblock=c("yr","mo"))
  mn_SEBS_season$basin = "SEBS"

  plot_data        <- rbind(mn_NEBS_season,mn_SEBS_season)

 # plot Fall values:
 p7 <- ggplot(data = plot_data%>%filter(mo==9) ) +
    geom_line(   aes(x = yr,y = mn_val,color=basin),alpha=.8)+
    geom_smooth( aes(x = yr,y = mn_val,color=basin),
                formula = y ~ x, se = T)+
    facet_grid(basin~.)+
    scale_color_viridis_d(begin=.4,end=.8)+
    ylab(tmp_var$units[1])+
    ggtitle( paste(aclim[2],"Sept.",mn_NEBS_season$var[1]))+
    theme_minimal()
p7

if(update.figs)
  ggsave(file=file.path(main,"Figs/Sept_large_Zoop.jpg"),width=8,height=5)
```
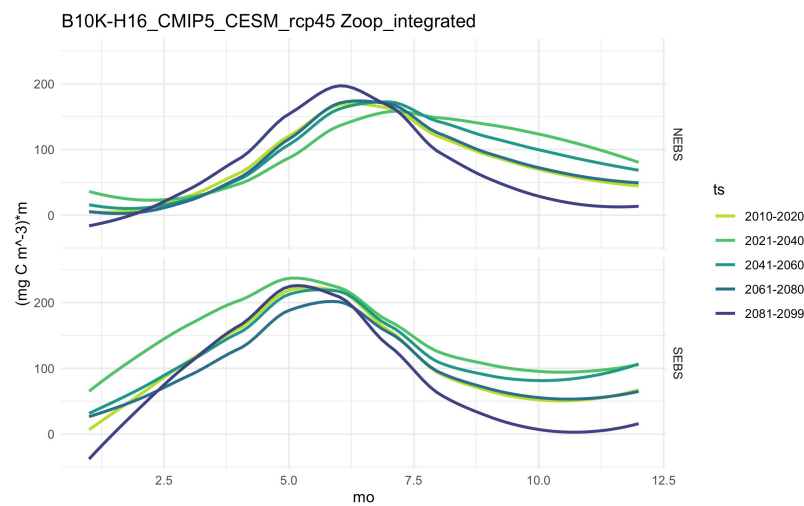

Figure 8: September large zooplankton integrated concentration

Finally we can use this approach to plot the monthly averages and look for phenological shifts:

```
 # or average in 4 time slices by mo:
 # 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))
```

```
  plot_data$ts <-names(time_seg)[1]
  for(tt in 1:length((time_seg)))
    plot_data$ts[plot_data$yr%in%(time_seg[[tt]][1]:time_seg[[tt]][2])]<-names(time_seg)[tt]

  plot_data2 <- plot_data%>%
    group_by(var,mo,units,long_name,basin, ts)%>%
    summarize(mn_val2 = mean(mn_val))

# now plot phenological shift:
 p8 <- ggplot(data = plot_data2 ) +
    geom_line(   aes(x = mo,y = mn_val2,color=ts),alpha=.8,size=0)+
    geom_smooth( aes(x = mo,y = mn_val2,color=ts),
                 formula = y ~ x, se = F)+
    facet_grid(basin~.)+
    scale_color_viridis_d(begin=.9,end=.2)+
    ylab(tmp_var$units[1])+
    ggtitle( paste(aclim[2],mn_NEBS_season$var[1]))+
    theme_minimal()
p8
if(update.figs)
  ggsave(file=file.path(main,"Figs/PhenShift_large_Zoop.jpg"),width=8,height=5)
```


Figure 9: September large zooplankton integrated concentration

4.2 Level 2:

Level 2 data can be explored in the same way as the above indices but we will focus in the section below on a simple spatial plot and temporal index. The advantage of Level2 inidces is in the spatial resolution and values outside of the survey area.

3.2.1 Explore Level 2 data catalog

3.2.2 Custom spatial indices

```r
# define 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))

# View an individual variable (e.g., Bottom Temp)
# ----------------------------------------------------
head(srvy_vars)
head(aclim)

# assign the simulation to download
# --> --> Tinker: try selecting a different set of models to compare
sim        <-"B10K-H16_CMIP5_MIROC_rcp85"

svl <- list(
   'Bottom 5m' = "temp",
   'Surface 5m' = "temp",
   'Integrated' = c("EupS","Cop","NCaS") )

# Currently available Level 2 variables
dl      <- proj_l2_datasets$dataset  # datasets


# Let's sample the model years as close to Aug 1 as the model timesteps run:
tr          <- c("-08-1 12:00:00 GMT")

# the full grid is large and takes a longtime to plot, so let's subsample the grid every 4 cells

IDin       <- "_Aug1_subgrid"
var_use    <- "_bottom5m_temp"

# open a "region" or strata specific nc file
fl          <- file.path(main,Rdata_path,sim,"Level2",paste0(sim,var_use,IDin,".Rdata"))

# load object 'ACLIMsurveyrep'
if(!file.exists(fl))
  get_l2(
    ID          = IDin,
    xi_rangeIN  = seq(1,182,10),
    eta_rangeIN = seq(1,258,10),
    ds_list     = dl,
    trIN        = tr,
    sub_varlist = svl,
    sim_list    = sim  )

# load R data file
load(fl)    # temp
```

```r
# there are smarter ways to do this;looping because
# we don't want to mess it up but this is slow...
i <-1
data_long <- data.frame(latitude = as.vector(temp$lat),
                        longitude = as.vector(temp$lon),
                        val = as.vector(temp$val[,,i]),
                        time = temp$time[i],
                        year = substr( temp$time[i],1,4)
                        )
for(i in 2:dim(temp$val)[3])
  data_long <- rbind(data_long,
                     data.frame(latitude = as.vector(temp$lat),
                       longitude = as.vector(temp$lon),
                       val = as.vector(temp$val[,,i]),
                       time = temp$time[i],
                       year = substr( temp$time[i],1,4))
                     )


# get the mean values for the time blocks from the rdata versions
# will throw "implicit NA" errors that can be ignored
tmp_var <-data_long # get mean var val for each time segment
j<-0
for(i in 1:length(time_seg)){
  if(length(which(tmp_var$year%in%time_seg[[i]]))>0){
    j <- j +1
     mn_tmp_var <- tmp_var%>%
      filter(year%in%time_seg[[i]],!is.na(val))%>%
      group_by(latitude, longitude)%>%
      summarise(mnval = mean(val,rm.na=T))

    mn_tmp_var$time_period = factor(names(time_seg)[i],levels=names(time_seg))
  if(j == 1) mn_var <- mn_tmp_var
  if(j >  1) mn_var <- rbind(mn_var,mn_tmp_var)
   rm(mn_tmp_var)
  }
}

# convert results to a shapefile
L2_sf  <- convert2shp(mn_var%>%filter(!is.na(mnval)))

p9      <- plot_stations_basemap(sfIN = L2_sf,
                            fillIN = "mnval",
                            colorIN = "mnval",
                            sizeIN=.6) +
  facet_grid(.~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(paste(sim,var_use,IDin))
```

```
# This is slow but it works (repeat dev.new() twice if in Rstudio)...
dev.new()
p9

if(update.figs)  ggsave(file=file.path(main,"Figs/sub_grid_mn_BT_Aug1.jpg"),width=8,height=6)

# graphics.off()
```
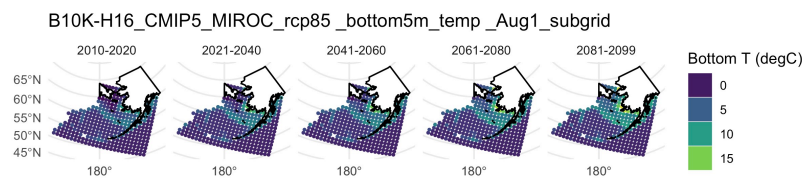

Figure 10: Aug 1 Bottom temperature from Level 2 dataset

6. 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:

7. Helpful links and further reading:

7.1 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.

7.2 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**

7.3 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.