pacea

an R package of Pacific ecosystem information to help facilitate an ecosystem approach to fisheries management

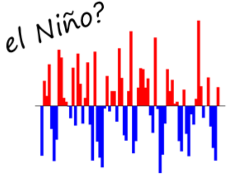
Andrew Edwards and Travis Tai

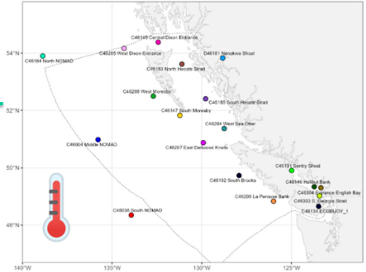
pacea: an R package of ecosystem information to help facilitate an ecosystem approach to fisheries management

Andrew Edwards[1,2] & Travis Tai[1]

[1]Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC.

[2]Dept. of Biology, University of Victoria, Victoria, BC.





National EAFMWG Webinar Series
9th May 2024

Fisheries and Oceans Canada Pêches et Océans Canada

Acknowledgments

- Co-authors: Joe Watson, Angelica Peña, Andrea Hilborn, Charles Hannah, and Chris Rooper.
- Kelsey Flynn, Jessica Nephin, Lindsay Davidson, Strahan Tucker, Brianna Wright, Patrick Thompson, Matt Grinnell, Sean Anderson, Philina English, Chris Grandin, Jennifer Boldt, and others.
- Carley Colclough for designing the logo.
- DFO's Competitive Science Research Fund for funding (project 21-FS-03-13).

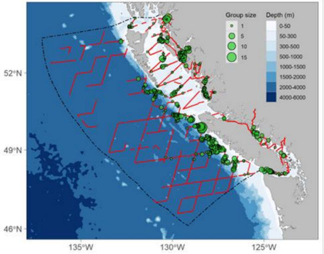
Motivation

- Revised Fisheries Act: "… the Minister shall take into account the environmental conditions affecting a fish stock."
- Yet $<$50% of DFO's stock assessments currently use environmental data.
- Only 28% of assessments in Pacific Region use environmental data.
- Leading cause of not using environmental data is availability of the data.

Kulka DW, Thompson S, Cogliati K, Olmstead M, Austin D, Pepin D. (2022). An Accounting of Integration of Environmental Variables in Fishery Stock Assessments in Canada. Can. Tech. Rep. Fish. Aquat. Sci. 3473: viii + 79 p.
https://publications.gc.ca/collections/collection_2022/mpo-dfo/Fs97-6-3473-eng.pdf

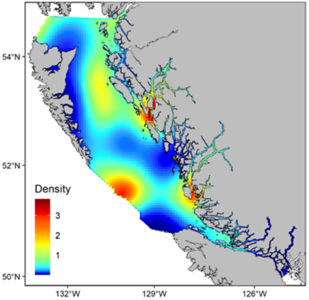
Motivation (based on a true story)

Survey sightings of Humpback Whales



Modelling

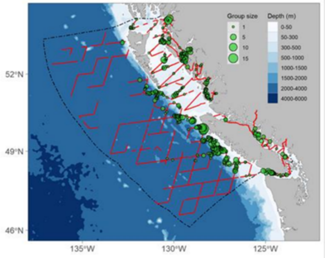
Estimated densities



See Doniol-Valcroze et al. (2022) in last year's SOPO report.
Density plot courtesy Brianna Wright.

Motivation (based on a true story)

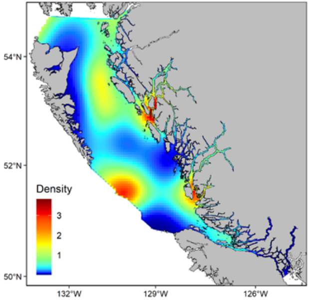
Survey sightings of Humpback Whales



Modelling

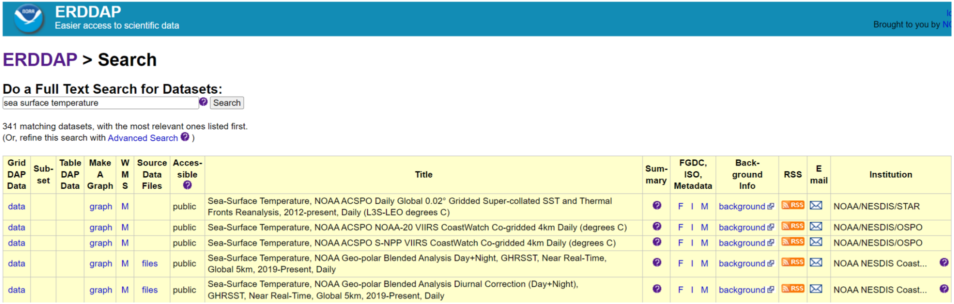
SST ?

Estimated densities



Motivation (based on a true story)

A search for sea surface temperature yields an overwhelming number (341) of choices:



ERDDAP
Easier access to scientific data
Brought to you by N

ERDDAP > Search

Do a Full Text Search for Datasets:
sea surface temperature | Search

341 matching datasets, with the most relevant ones listed first.
(Or, refine this search with Advanced Search)

Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	W M S	Source Data Files	Acces-sible	Title	Sum-mary	FGDC, ISO, Metadata	Back-ground Info	RSS	E mail	Institution
data			graph	M		public	Sea-Surface Temperature, NOAA ACSPO Daily Global 0.02° Gridded Super-collated SST and Thermal Fronts Reanalysis, 2012-present, Daily (L3S-LEO degrees C)		F I M	background	RSS		NOAA/NESDIS/STAR
data			graph	M		public	Sea-Surface Temperature, NOAA ACSPO NOAA-20 VIIRS CoastWatch Co-gridded 4km Daily (degrees C)		F I M	background	RSS		NOAA/NESDIS/OSPO
data			graph	M		public	Sea-Surface Temperature, NOAA ACSPO S-NPP VIIRS CoastWatch Co-gridded 4km Daily (degrees C)		F I M	background	RSS		NOAA/NESDIS/OSPO
data			graph	M	files	public	Sea-Surface Temperature, NOAA Geo-polar Blended Analysis Day+Night, GHRSST, Near Real-Time, Global 5km, 2019-Present, Daily		F I M	background	RSS		NOAA NESDIS Coast...
data			graph	M	files	public	Sea-Surface Temperature, NOAA Geo-polar Blended Analysis Diurnal Correction (Day+Night), GHRSST, Near Real-Time, Global 5km, 2019-Present, Daily		F I M	background	RSS		NOAA NESDIS Coast...

Likely requires extensive data wrangling to be usable, which usually takes way, way, longer than anticipated.
So the SST analysis did not happen.

Motivation

- Initial motivation was Dan Duplisea's `gslea` package for the Gulf of St. Lawrence.
- Outputs from a BC physical biogeochemical model have been shared by Angelica Peña.
- But the netCDF format may be unfamiliar to non-oceanographers – we take care of the wrangling into R.
- Open Data is great, but can be hard to convert raw data into usable information.
- Primary audience is DFO stock assessment scientists, but usable by anyone (with a minimal working knowledge of R).

Duplisea et al. (2020). gslea: the Gulf of St Lawrence ecosystem approach data matrix R-package. R package version 0.1 https://github.com/duplisea/gslea

Why an R package?

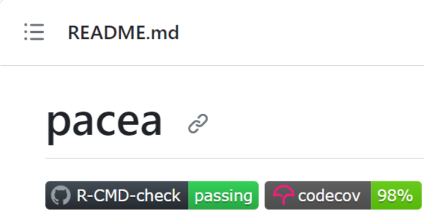
- R is the computer programming language most widely used by stock assessment scientists.
- An R package is the standard way to share code (and data).
- Easy installation:

```
remotes::install_github("pbs-assess/pacea")
```

- Ensures proper documentation of data objects and functions (helps users).
- Include vignettes that walk users through various features.
- All data (except one type) is saved within the package, no further downloading required, so not relying on external websites being functional.

Why an R package on GitHub?

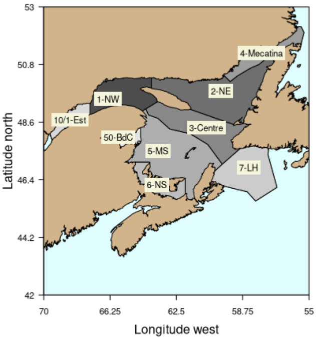
- Can host freely on GitHub: https://github.com/pbs-assess/pacea
- Standard way to collaborate on code.
- Code is completely open source: transparent, traceable, and transferable.



- 'Passing' badge: ensures package can build in R.
- 'codecov' is the amount of the code that is covered by unit tests, which help weed out bugs.

Spatial domain

Started thinking about ecosystem assessment regions like in gslea:



But it became non-trivial to decide on regions, and we decided to explicitly include spatial information, especially regarding outputs of the biogeochemical model.

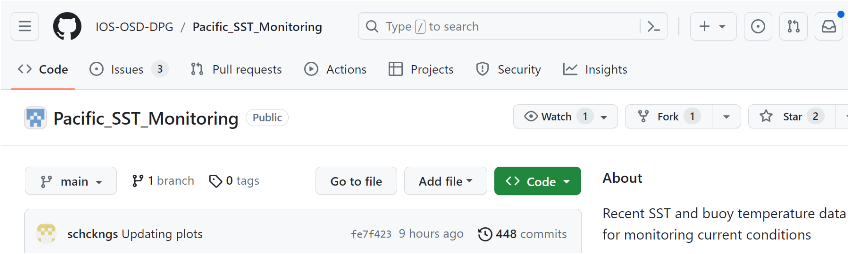
pacea objects

Currently, pacea contains:

- 205,179 calculations of daily sea surface temperature based on data from 19 buoys
- outputs from the spatial British Columbia continental margin (BCCM) model
- NOAA's spatial Optimum Interpolation Sea Surface Temperature (OISST) record
- 9 climatic and oceanographic indices, such as the Pacific Decadal Oscillation and those related to El Niño
- estimates of abundances for Harbour Seals
- estimates of spawning stock biomass and annual recruitments for Pacific Hake and Pacific Herring
- zooplankton biomass anomalies

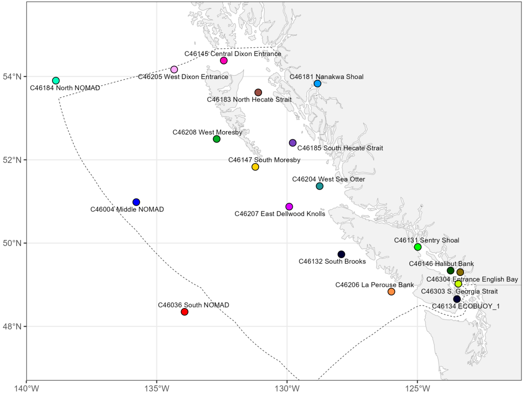
Sea surface temperature

Andrea Hilborn (IOS) maintains https://github.com/IOS-OSD-DPG/Pacific_SST_Monitoring which is updated weekly to show recent SST from various sources for the Northeast Pacific (and anomalies and marine heatwave information).



Sea surface temperature data from buoys

Data from 19 buoys maintained by DFO/ECCC (Environment and Climate Change Canada), hosted through CIOOS (Canadian Integrated Ocean Observing System).



Sea surface temperature data from buoys

Extensive data processing to calculate daily mean SST:

- Resolution is every 10 minutes or every hour
- Come with quality control flags (mostly)
- We decided that for a daily mean we require:
    - at least one measurement every 2 hours
    - $\geq$ 10 2-hour measurements every day
- Switch from DFO data stream to ECCC data stream when appropriate
- Time zones and daylight savings time
- Still working on fixing a few outliers in the high-resolution data (hard to automate)

Sea surface temperature data from buoys

The data are saved as a `tibble` in `pacea`:

```
buoy_sst
# A tibble: 205,179 x 3
  date       stn_id   sst
  <date>     <fct>  <dbl>
1 1988-08-05 C46004  12.8
2 1988-08-06 C46004  12.7
3 1988-08-07 C46004  12.5
4 1988-08-08 C46004  12.5
5 1988-08-09 C46004  12.6
6 1988-08-10 C46004  12.6
# i 205,173 more rows
```

Sea surface temperature data from buoys

The data are saved as a `tibble` in `pacea`:

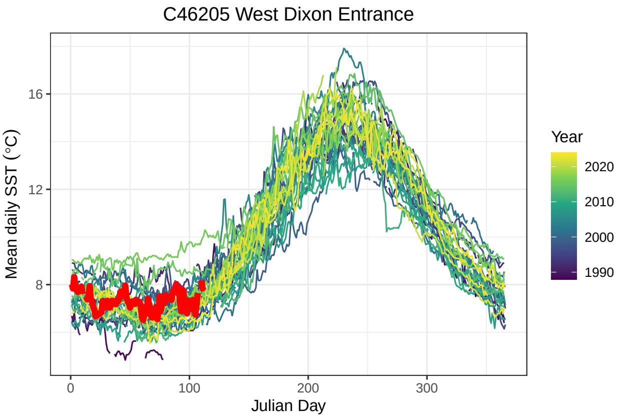
```
buoy_sst
# A tibble: 205,179 x 3
  date       stn_id   sst
  <date>     <fct>  <dbl>
1 1988-08-05 C46004  12.8
2 1988-08-06 C46004  12.7
3 1988-08-07 C46004  12.5
4 1988-08-08 C46004  12.5
5 1988-08-09 C46004  12.6
6 1988-08-10 C46004  12.6
# i 205,173 more rows
```

```
tail(buoy_sst)
# A tibble: 6 x 3
  date       stn_id   sst
  <date>     <fct>  <dbl>
1 2024-04-16 C46304  9.57
2 2024-04-17 C46304 10.0
3 2024-04-18 C46304 10.1
4 2024-04-19 C46304 10.1
5 2024-04-20 C46304  9.92
6 2024-04-21 C46304 10.6
```

Sea surface temperature data from buoys

Plot data from a single buoy for all years, the default is for buoy C46205:

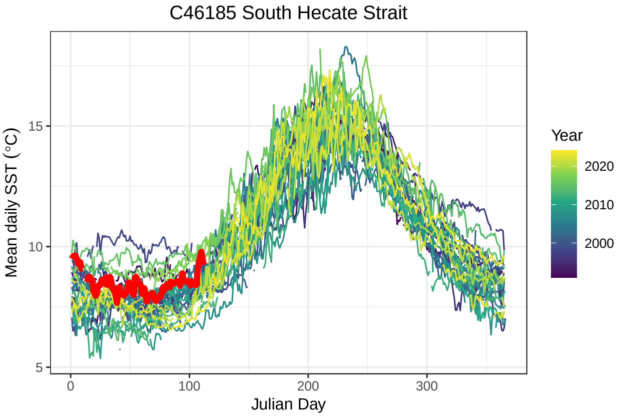
```
plot(buoy_sst)
```



Sea surface temperature data from buoys

Easy to look at values for any buoy

```
plot(buoy_sst,
     stn_id = "C46185")
```



Help file for every object

```
?buoy_sst

Daily average sea surface temperatures for 19 buoys (yielding over
200,000 values) calculated with data from Environment and Climate
Change Canada (ECCC) and DFO.

Description:

     A tibble of daily average calculations of sea surface temperature
     in coastal Canadian Pacific waters. The earliest data are from
     September 1987, and 14 buoys were still providing data as of May
     2023. See the example code below to see the start and end dates
     for each buoy.

Usage:

     buoy_sst

Format:

     A tibble with columns:

     date: date (of class `Date`) used to calculate the `sst`, based on
          UTC -8 hours (i.e. Pacific Standard Time, not changing due to
          daylight savings)

    ...
```

Vignettes for each type of data

We have written vignettes for each type of data. Viewable directly on GitHub site (see the README).

Buoy Sea Surface Temperature Data

Andrew Edwards

Daily average sea surface temperature from buoys

```
library(pacea)
library(dplyr)
library(tibble)  # Else prints all of a tibble
```

In pacea we include daily average sea surface temperature (SST) from 19 buoys in Canadian Pacific waters, yielding over 170,000 values. Data are from Environment and Climate Change Canada, and Fisheries and Oceans Canada. The earliest data are from September 1987, and 14 buoys were still providing data as of May 2023.

Metadata for the buoys is given by

```
buoy_metadata
#> # A tibble: 19 × 9
#>    wmo_id name     type  latitude longitude water_depth_m col_key stn_id name_key
#>    <fct>  <fct>    <fct>    <dbl>     <dbl>         <dbl> <fct>   <fct>  <fct>
#>  1 46004  Middle… NOMAD     51.0      -136.          3600 #0000FF C46004 C46004 …
#>  2 46036  South … NOMAD     48.4      -134.          3500 #FF0000 C46036 C46036 …
#>  3 46131  Sentry… 3 me…     49.9      -125.            18 #00FF00 C46131 C46131 …
#>  4 46132  South … 3 me…     49.7      -128.          2040 #000033 C46132 C46132 …
#>  5 46134  ECOBUO… 3 me…     48.7      -123.            65 #000033 C46134 C46134 …
#>  6 46145  Centra… 3 me…     54.4      -132.           257 #FF00B6 C46145 C46145 …
```

Climatic and oceanographic indices

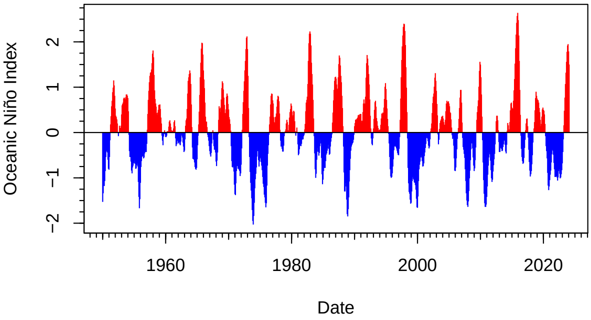
Various climate and oceanographic indices are currently included in pacea:

Object	Description	Resolution	Start year	End year
pdo	Pacific Decadal Oscillation	monthly	1854	2024
npi_monthly	North Pacific Index (monthly)	monthly	1899	2024
npi_annual	North Pacific Index (annual)	annual	1899	2024
alpi	Aleutian Low Pressure Index	annual	1900	2022
oni	Oceanic Niño Index	monthly	1950	2024
npgo	North Pacific Gyre Oscillation	monthly	1950	2024
ao	Arctic Oscillation	monthly	1950	2024
soi	Southern Oscillation Index	monthly	1951	2024
mei	Multivariate El Niño Southern Oscillation Index	monthly	1979	2024

Climatic and oceanographic indices

For example, to see a plot of the Oceanic Niño Index (ONI) anomaly, simply type

```
plot(oni)
```



This shows the onset of El Niño (positive index) conditions in April 2023.

Climatic and oceanographic indices

The values are readily available for doing your own analyses:

```
oni
# A tibble: 890 x 4
   year month value anomaly
  <dbl> <dbl> <dbl>   <dbl>
1  1950     1  24.7   -1.53
2  1950     2  25.2   -1.34
3  1950     3  25.8   -1.16
4  1950     4  26.1   -1.18
5  1950     5  26.3   -1.07
6  1950     6  26.3   -0.85
# i 884 more rows
tail(oni)
# A tibble: 6 x 4
   year month value anomaly
  <dbl> <dbl> <dbl>   <dbl>
1  2023     9  28.3    1.56
2  2023    10  28.5    1.78
3  2023    11  28.6    1.92
4  2023    12  28.6    1.95
5  2024     1  28.4    1.79
6  2024     2  28.4    1.5
```

Climatic and oceanographic indices

Each climatic and oceanographic index is saved as a data object, and properly documented, described, and referenced in its help file, for example:

```
?oni

ONI - Oceanographic Niño Index

Description:

     The Oceanic Niño Index is a monthly index which is one measure of
     the El Niño-Southern Oscillation.

Usage:

     oni

Format:

     A tibble also of class `pacea_index` with columns:

     year: year of value

     month: month (1 to 12) of value

     val: absolute values of three-month averages (preceding, current,
          and next month), deg C; note that recent values may change in
          subsequent updates - see details

     anom: anomalies based on 30-year base periods that are updated
```

Climatic and oceanographic indices

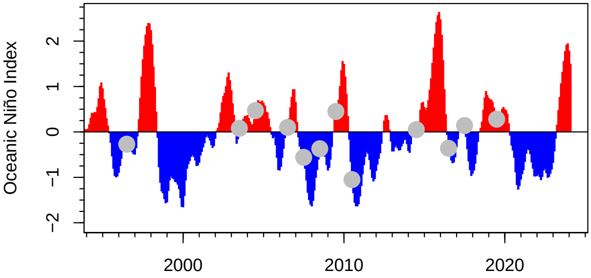
Continuing the help. . .

```
The Oceanic Niño Index (ONI) is a 3-month running mean of sea
surface temperature (SST) anomalies in the Niño 3.4 region (5 deg
N to 5 deg S, 120 deg W to 170 deg W) plotted on the center month.
The SST anomalies are calculated based on 30-year base periods
that are updated every 5 years, which accounts for global warming
and some of the decadal-scale SST variability (as seen in the
Pacific Decadal Oscillation index). The ONI is provided by the
NOAA's National Weather Service National Centers for Environmental
Prediction CPC:
http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml

...

    Because of the high frequency filter applied to the ERSSTv5 data,
`ONI values may change up to two months after the initial real
time value is posted`. Therefore, the most recent ONI values
should be considered an estimate.
```

Further plotting options and styles

Say you want to see if specific events coincide with El Niño (based on a true story, see help for details):

```
plot(oni,
     event_years = c(1996, 2003, 2004, 2006, 2007, 2008, 2009, 2010, 2014, 2016, 2017, 2019),
     xlim = c(lubridate::dmy(01011995),
              lubridate::dmy(01012024)),
     lwd = 2)
```



Estimates of animal populations

In pacea we include estimates of populations as calculated from recent stock assessments or analyses.

Currently, we have:

- Pacific Hake spawning biomass and recruitments for the coastwide stock off Canada and United States
- Pacific Harbour Seals abundance (the most abundant pinniped species in the Northeast Pacific)
- Pacific Herring spawning biomass and recruitments for all five stock assessment regions
- Zooplankton biomass anomalies for 24 species groups in the Strait of Georgia

Hake biomass

The hake biomass time series is saved as a `tibble` object in `pacea`:

```
hake_biomass
# A tibble: 59 x 4
   year   low median  high
  <dbl> <dbl>  <dbl> <dbl>
1  1966 0.576  0.954  1.82
2  1967 0.587  0.959  1.83
3  1968 0.590  0.967  1.92
4  1969 0.685  1.10   2.19
5  1970 0.773  1.26   2.53
6  1971 0.785  1.30   2.62
# i 53 more rows

tail(hake_biomass)
# A tibble: 6 x 4
   year   low median  high
  <dbl> <dbl>  <dbl> <dbl>
1  2019 1.00    1.40  2.50
2  2020 0.911   1.35  2.57
3  2021 0.699   1.12  2.27
4  2022 0.627   1.12  2.45
5  2023 0.652   1.34  3.22
6  2024 0.853   1.88  4.83
```

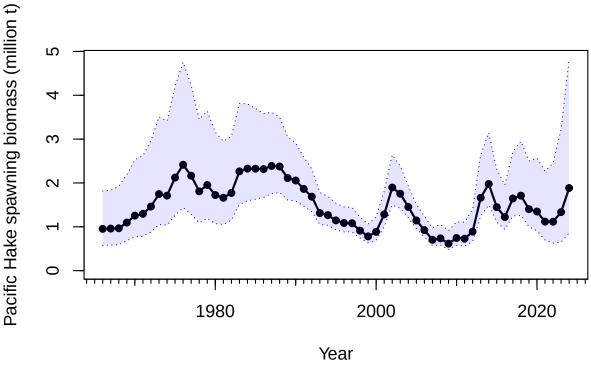
Hake biomass

```
?hake_biomass

Pacific Hake annual spawning stock biomass (females only) as estimated
by the 2024 stock assessment.

Description:

     The Pacific Hake stock is managed and assessed through an
     Agreement between Canada and the United States. The spawning stock
     biomass (mature females) estimates come from the most recent joint
     stock assessment, and, importantly, are for the coastwide stock
     from California to British Columbia.

Usage:

     hake_biomass

Format:

     A tibble also of class `pacea_biomass` with columns:

     year: year of the estimate of spawning stock biomass (mature
          females), where the estimate is for the start of the year

     low: low end (2.5th percentile) of the 95% credible interval for
          biomass, units as for `median`

     median: median estimate of biomass, in millions of tonnes of
          female spawning biomass at the start of the year
```

Hake biomass

Time series of spawning biomass:

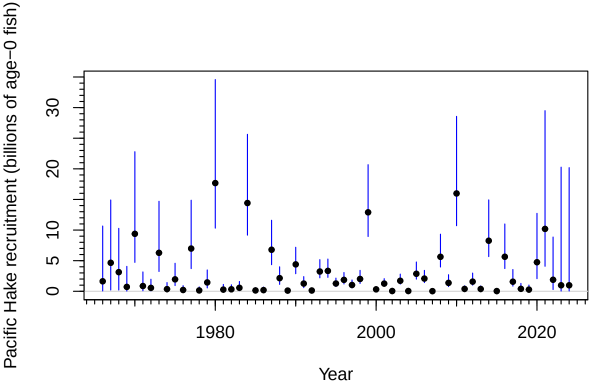
```
plot(hake_biomass)
```



Hake recruitment

Estimates of annual recruitment (exclude recent years for any analyses since not greatly informed by data; see `?hake_recruitment`):

```
plot(hake_recruitment)
```



Herring biomass

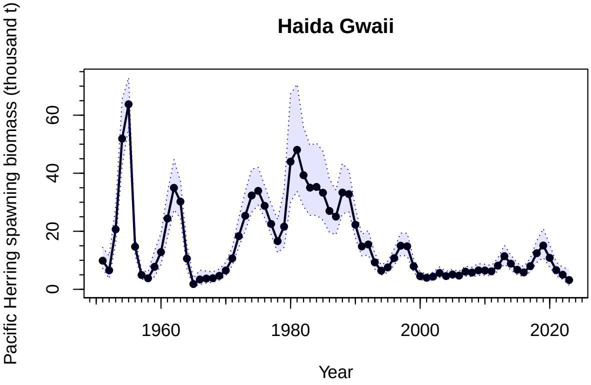
The herring spawning biomass time series is saved similarly to hake, but contains values for five regions:

```
herring_spawning_biomass
# A tibble: 365 x 5
   year region   low median  high
  <dbl> <fct>  <dbl>  <dbl> <dbl>
1  1951 HG      7.20   9.86  14.4
2  1952 HG      3.71   6.55  11.5
3  1953 HG     15.8   20.7   28.6
4  1954 HG     42.7   51.9   65.5
5  1955 HG     55.9   63.8   72.9
6  1956 HG     11.7   14.7   18.0
# i 359 more rows

tail(herring_spawning_biomass)
# A tibble: 6 x 5
   year region   low median  high
  <dbl> <fct>  <dbl>  <dbl> <dbl>
1  2018 WCVI    11.8   15.3  19.6
2  2019 WCVI    11.7   15.4  20.2
3  2020 WCVI    13.7   18.1  24.2
4  2021 WCVI    17.8   24.0  32.8
5  2022 WCVI    24.2   34.7  49.4
6  2023 WCVI    24.4   41.2  67.5
```

Herring biomass

Estimates of spawning biomass for one region (HG = Haida Gwaii):

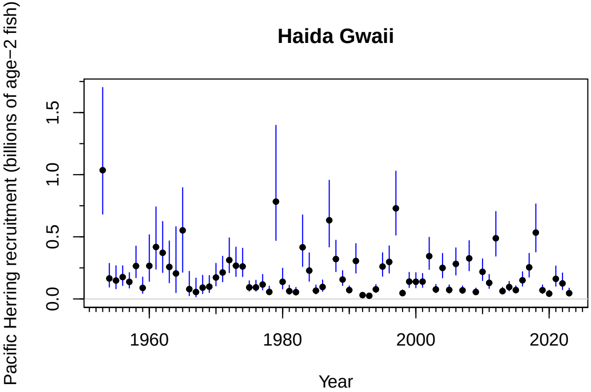
```
plot(herring_spawning_biomass,
     region = "HG")
```



Herring recruitment

Estimates of annual age-2 recruitment for one region (HG = Haida Gwaii):

```
plot(herring_recruitment,
     region = "HG")
```



Pacific Harbour Seals

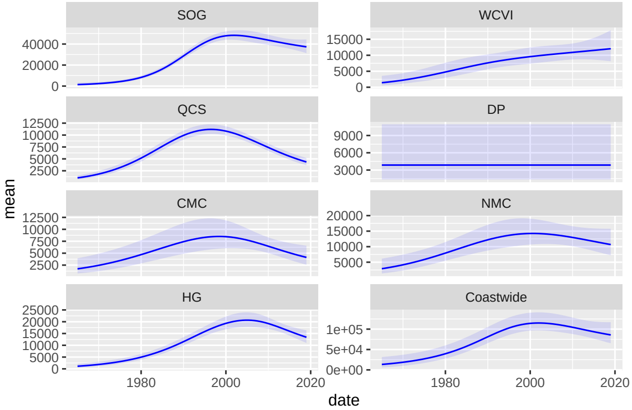
Estimated abundances for seven regions (and coastwide) were calculated by DFO (2022; SAR 2022/034), and included here:

```
harbour_seals
# A tibble: 640 x 5
  date       region   low  mean  high
  <date>     <fct>  <dbl> <dbl> <dbl>
1 1965-01-01 SOG      593. 1384. 3231.
2 1965-09-07 SOG      669. 1488. 3307.
3 1966-05-14 SOG      755. 1599. 3387.
4 1967-01-19 SOG      851. 1719. 3471.
5 1967-09-25 SOG      959. 1848. 3562.
6 1968-06-01 SOG     1080. 1988. 3661.
# i 634 more rows

tail(harbour_seals)
# A tibble: 6 x 5
  date       region       low  mean    high
  <date>     <fct>      <dbl> <dbl>   <dbl>
1 2015-08-01 Coastwide 75072. 92401. 118166.
2 2016-04-07 Coastwide 73146. 91001. 117663.
3 2016-12-13 Coastwide 71189. 89638. 117317.
4 2017-08-20 Coastwide 69222. 88313. 117113.
5 2018-04-26 Coastwide 67259. 87026. 117031.
6 2019-01-01 Coastwide 65316. 85774. 117053.
```

Pacific Harbour Seals

Estimates for each region and coastwide, reproducing Figure 3 of DFO (2022):

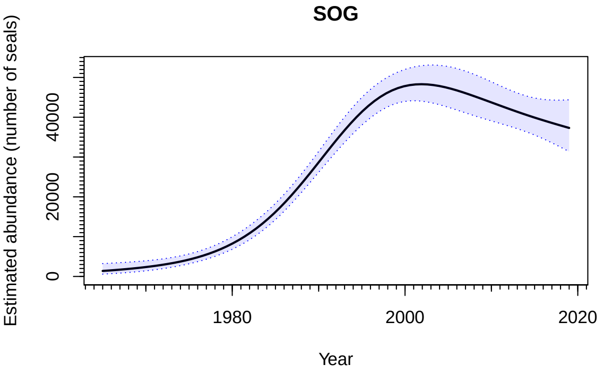
```
plot(harbour_seals)
```



Pacific Harbour Seals

Or just plot estimates for a single region:

```
plot(harbour_seals,
     region = "SOG")
```



Zooplankton in the Strait of Georgia

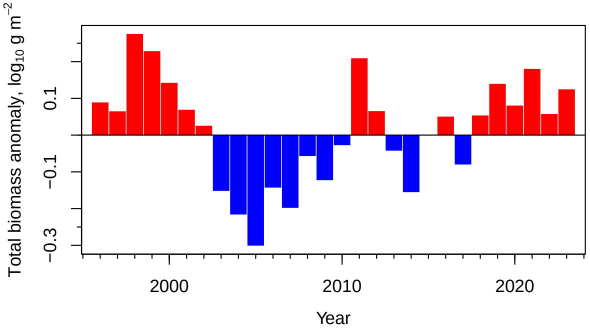
Biomass anomalies of zooplankton biomass from 1996 onwards from Perry et al. (2021), as extended by Kelly Young each year for DFO's State of the Pacific Ocean Report.

```
zooplankton_sog
# A tibble: 28 x 28
   year number_samples volume_filtered total_biomass amphipods_gammarid
  <dbl>          <dbl>           <dbl>         <dbl>              <dbl>
1  1996              4            335.        0.0890             0.0338
2  1997             11            921.        0.0649             0.212
3  1998             22           1584.        0.275              0.184
4  1999              7            421.        0.229              0.0139
5  2000              4            484.        0.142              0.212
6  2001              8            430.        0.0690            -0.0154
# i 22 more rows
# i 23 more variables: amphipods_hyperiid <dbl>, benthic_larvae <dbl>,
#   calanoid_copepods_large <dbl>, calanoid_copepods_medium <dbl>,
#   calanoid_copepods_small <dbl>, cephalopoda <dbl>, chaetognatha <dbl>,
#   cladocera <dbl>, ctenophora <dbl>, euphausiids <dbl>, fish <dbl>,
#   larvacea <dbl>, medusae <dbl>, mysids <dbl>, natantia <dbl>,
#   non_calanoid_copeopods <dbl>, ostracoda <dbl>, other <dbl>, ...
```

Zooplankton in the Strait of Georgia

Default plot of anomalies of total biomass:

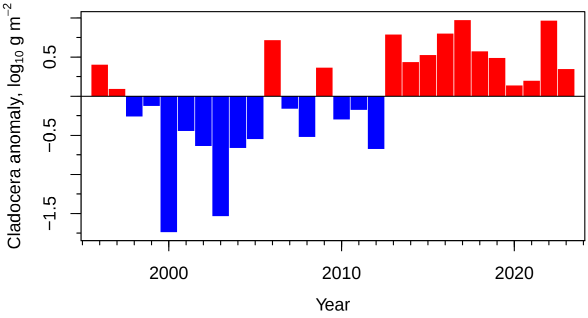
```
plot(zooplankton_sog)
```



Zooplankton in the Strait of Georgia

Or plot anomalies for a specific species group:

```
plot(zooplankton_sog,
     species_group = "cladocera")
```



NOAA's Optimal Interpolation Sea Surface Temperature

OISST - long term gridded climate data record

- $1/4°$ longitude x latitude
- Interpolated to fill gaps
- Canada's Pacific EEZ
- 925 gridded cells
- September, 1981 to (relatively) present
    - Updated ~monthly

2 data products: weekly and monthly mean SST

- `oisst_7day` and `oisst_month`

R `erddap` - https://www.ncei.noaa.gov/products/optimum-interpolation-sst

Huang, B., C. Liu, V. Banzon, E. Freeman, G. Graham, B. Hankins, T. Smith, and H. Zhang, 2021: Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1. J. Climate, 34, 2923–2939, https://doi.org/10.1175/JCLI-D-20-0166.1.

OISST – monthly mean

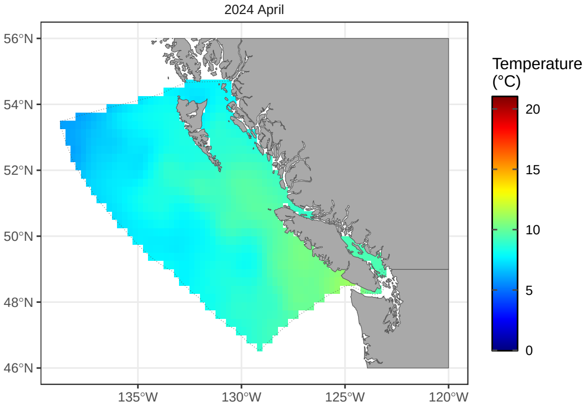
The data are saved as a `sf` (simple feature) `tibble` object in `pacea`:

```
oisst_month
```

```
Simple feature collection with 467125 features and 7 fields
Geometry type: POINT
Dimension:     XY
Bounding box:  xmin: -138.625 ymin: 46.625 xmax: -123.125 ymax: 54.625
Geodetic CRS:  WGS 84
# A tibble: 467,125 x 8
   year month   sst sst_sd sst_n start_date end_date
* <dbl> <dbl> <dbl>  <dbl> <int> <date>     <date>
1  1981     9  16.4  0.830    30 1981-09-01 1981-09-30
2  1981    10  14.3  0.838    31 1981-10-01 1981-10-31
3  1981    11  12.7  1.38     30 1981-11-01 1981-11-30
4  1981    12  10.2  0.549    31 1981-12-01 1981-12-31
5  1981     9  16.4  1.05     30 1981-09-01 1981-09-30
6  1981    10  14.2  0.697    31 1981-10-01 1981-10-31
# i 467,119 more rows
# i 1 more variable: geometry <POINT [B0]>
# i Use `print(n = ...)` to see more rows
```

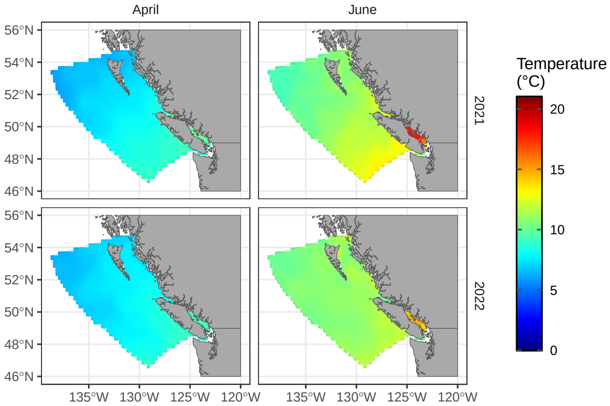
Built in plots – default

```
plot(oisst_month)
```



Built in plots – select years and months

```
plot(oisst_month, months.plot = c("April", "June"), years.plot = c(2021:2022))
```



British Columbia Continental Margin (BCCM) model

Physical biogeochemical oceanographic model

- Regional Ocean Modelling Systen (ROMS)
- Curvilinear grid at 3km x 3km resolution
- Interpolated to regular grid
- Clipped to Canada's Pacific EEZ

Model output provided by Angelica Peña at the Institute of Ocean Sciences (Fisheries and Oceans Canada).

Peña, M.A., Fine, I. and Callendar, W. 2019. Interannual variability in primary production and shelf-offshore transport of nutrients along the northeast Pacific Ocean margin. Deep-Sea Research II, doi:10.1016/j.dsr2.2019.104637.

Nearest neighbour linear interpolation

Curvilinear –> regular grid

- Inshore (2 x 2km) and offshore (6 x 6km)
- Nearest neighbour interpolation
    - Performace = RMSE error
- 40,480 cells
- Monthly from 1993-2019

Available variables

The variables are:

- dissolved oxygen concentration
- pH
- salinity
- temperature
- depth-integrated phytoplankton
- depth-integrated primary production.

For applicable variables these are given for

- sea surface
- 0-40 m integration
- 40-100 m integration
- 100 m to the sea bottom
- sea bottom.

List of available variables

bccm_data for available variables

```
bccm_data
                           data_name
1             bccm_surface_oxygen
2                 bccm_surface_ph
3           bccm_surface_salinity
4        bccm_surface_temperature
5           bccm_avg0to40m_oxygen
6               bccm_avg0to40m_ph
7         bccm_avg0to40m_salinity
8      bccm_avg0to40m_temperature
9         bccm_avg40to100m_oxygen
10            bccm_avg40to100m_ph
11      bccm_avg40to100m_salinity
12   bccm_avg40to100m_temperature
13       bccm_avg100mtoBot_oxygen
14           bccm_avg100mtoBot_ph
15     bccm_avg100mtoBot_salinity
16  bccm_avg100mtoBot_temperature
17             bccm_bottom_oxygen
18                 bccm_bottom_ph
19           bccm_bottom_salinity
20        bccm_bottom_temperature
21            bccm_phytoplankton
22        bccm_primaryproduction
```

Downloading data locally

The data are saved as `sf` (simple feature) objects in a separate GitHub repository `pacea-data` (https://github.com/pbs-assess/pacea-data).

One-time download to local machine (will take several minutes):

```
bccm_all_variables()

Downloading all BCCM data may take many minutes. Files will be downloaded to pacea_cache directory:
C:\Users\TAIT\AppData\Local/pacea/Cache
Would you like to continue?
1: Yes
2: No

Selection:
```

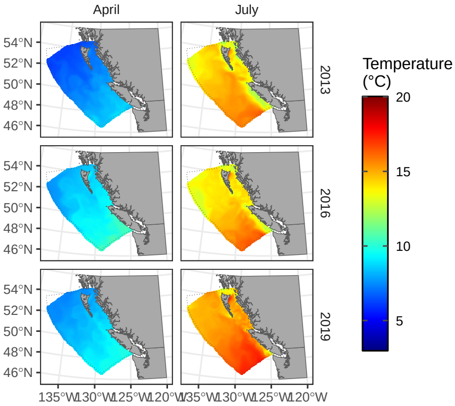
Example information

Average of pH from 100 m to sea bottom:

```
> bccm_avg100mtoBot_ph()

Simple feature collection with 33642 features and 324 fields
Geometry type: POLYGON
Dimension:     XY
Bounding box:  xmin: 164671.6 ymin: 163875 xmax: 1097612 ymax: 1103653
Projected CRS: NAD83 / BC Albers
# A tibble: 33,642 × 325
  `1993_1` `1993_2` `1993_3` `1993_4` ...                        geometry
     <dbl>    <dbl>    <dbl>    <dbl> ...                   <POLYGON [m]>
1     7.83     7.87     7.88     7.90 ... ((575612.5 1103653, 577612.5 ~
2     7.83     7.87     7.88     7.90 ... ((577612.5 1103653, 579612.5 ~
3     7.83     7.87     7.88     7.90 ... ((579612.5 1103653, 581612.5 ~
4     7.83     7.87     7.88     7.89 ... ((581612.5 1103653, 583612.5 ~
5     7.79     7.82     7.84     7.82 ... ((619612.5 1103653, 621612.5 ~
6     7.79     7.82     7.83     7.82 ... ((621612.5 1103653, 623612.5 ~
# i 33,636 more rows
# i 320 more variables: ...
```

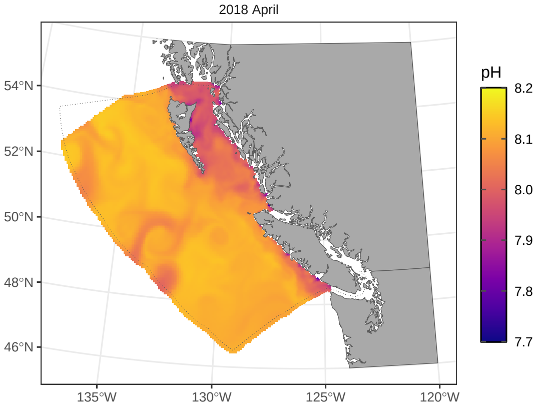
Built in plotting

```
plot(bccm_surface_temperature(), months.plot = c("april", "july"), years.plot = c(2013, 2016, 2019))
```



Built in plotting

```
plot(bccm_surface_ph())
```



Extras – climatology function

Built in functions to calculate 30-year climatology

```
?calc_clim
```

```
Description
  Function for calculating climatology of a pacea data object.

Usage
calc_clim(
  data,
  clim_years = c(1991:2020),
  clim_time = "month",
  time_period_return = "all"
)

Arguments
  data                 pacea data object: BCCM, OISST, buoy_sst data only

  clim_years             climatology period years

  clim_time            time units (e.g. month) to summarize climatologies

  time_period_return     vector of value(s) for the specific time units to estimate climatologies (e.g. '4' for week 4 or April). Set
...
```

Extras – climatology function

Output with each climatology time unit (e.g. week, month)

```
calc_clim(bccm_surface_temperature())
```

```
Simple feature collection with 486960 features and 4 fields
Geometry type: POLYGON
Dimension:     XY
Bounding box:  xmin: 164671.6 ymin: 163875 xmax: 1097612 ymax: 1103653
Projected CRS: NAD83 / BC Albers
# A tibble: 486,960 x 5
  month clim_value clim_sd clim_n                    geometry
  <dbl>      <dbl>   <dbl>  <int>               <POLYGON [m]>
1     1       5.80   0.721     27 ((575612.5 1103653, 5776~
2     2       5.47   0.610     27 ((575612.5 1103653, 5776~
3     3       5.64   0.627     27 ((575612.5 1103653, 5776~
4     4       6.61   0.623     27 ((575612.5 1103653, 5776~
5     5       8.49   0.732     27 ((575612.5 1103653, 5776~
6     6      10.5    0.662     27 ((575612.5 1103653, 5776~
# i 486,954 more rows
# i Use `print(n = ...)` to see more rows
```

And we can also use this with the OISST and buoy data

```
calc_clim(oisst_7day)
calc_clim(buoy_sst)
```

Extras – anomaly function

Built in functions to calculate anomaly to the 30 year climatology

```
?calc_anom
```

```
Description
  Function for calculating climatology of a pacea data object.

Usage
calc_anom(
  data,
  clim_years = c(1991:2020),
  clim_time = "month",
  time_period_return = "all",
  years_return
)

Arguments
  data                 pacea data object: BCCM, OISST, buoy_sst data only

  clim_year            climatology period years

  clim_time            time units (e.g. month) to summarize climatologies

  time_period_return     vector of value(s) for the specific time units to estimate climatologies (e.g. '4' for week 4 or April). Set t

  years_return         vector of value(s) to return the years of interest. Defaults to all years in input data
...
```

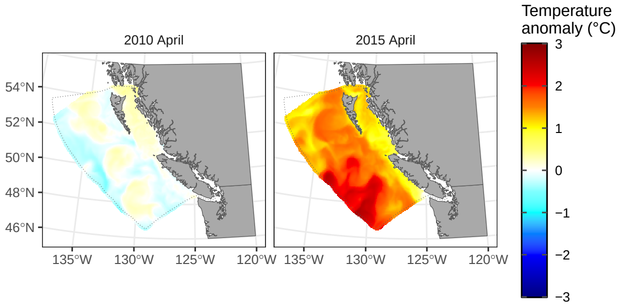
Extras – anomaly function

Specify year(s) and time period (e.g. week, month) to return – otherwise all.

```
anom_dat <- calc_anom(bccm_surface_temperature(), time_period_return = "April", years_return = c(2010, 2015))

Simple feature collection with 40580 features and 2 fields
Geometry type: POLYGON
Dimension:     XY
Bounding box:  xmin: 164671.6 ymin: 163875 xmax: 1097612 ymax: 1103653
Projected CRS: NAD83 / BC Albers
# A tibble: 40,580 x 3
  `2010_4` `2015_4`                                                                        geometry
     <dbl>    <dbl>                                                                   <POLYGON [m]>
1    0.114     1.09 ((575612.5 1103653, 577612.5 1103653, 577612.5 1101653, 575612.5 1101653, 575~
2    0.124     1.08 ((577612.5 1103653, 579612.5 1103653, 579612.5 1101653, 577612.5 1101653, 577~
3    0.150     1.07 ((579612.5 1103653, 581612.5 1103653, 581612.5 1101653, 579612.5 1101653, 579~
4    0.204     1.08 ((581612.5 1103653, 583612.5 1103653, 583612.5 1101653, 581612.5 1101653, 581~
5    0.204     1.07 ((583612.5 1103653, 585612.5 1103653, 585612.5 1101653, 583612.5 1101653, 583~
6    0.244     1.08 ((585612.5 1103653, 587612.5 1103653, 587612.5 1101653, 585612.5 1101653, 585~
# i 40,574 more rows
# i Use `print(n = ...)` to see more rows
```

Extras – anomaly plot

```
plot(anom_dat, months.plot = "April", years.plot = c(2010, 2015))
```



Extras – other tidbits in vignettes

Geospatial functions in vignettes

- `area_mean()` estimates inshore and offshore spatial mean (200m isobath)
- Masking/clipping specific areas (e.g. fishing regions of interest)
- Various other data visualisation examples

Commitments from package authors

We will:

- (try to) not break the package, by working on branches on GitHub before merging.
- ensure any improvements are back compatible.
- update the NEWS as we add/update data and make improvements.
- ideally update the following on the 20th of every month (or soon after) to get the latest raw data:
    - climatic and oceanographic indices
    - buoy sst
    - OISST

Commitments from package authors

- update other data as appropriate:
  - hake assessment output: February each year
  - harbour seals: next time assessment is updated
  - herring assessment output: when SAR published each year
  - BCCM model output: the model may be run into 2020 (boundary conditions come from a global model currently only available to Jan 2021).
    - note that BCCM is a research model (not an operational model) so updates depend on projects and access to High Performance Computing.

Recommended guidelines for users

Please:

- read the README!!
- check the help files and vignettes
- check the NEWS (for updates to data, useful additions, new data, etc.)
- understand the limitations of the data
- cite pacea and the original sources for any data you use

Current uses of pacea (that we know about)

- Spiny Dogfish: understanding declines over the past 20 years
- shrimp: impact of including environmental variables on predicting distributions
- humpback whales and porpoises: environmental covariates in species distribution models
- Pacific Saury assessment: finding links between two basin-scale indices and process errors, directly supporting Canada's North Pacific Fisheries Commission commitments

Lessons learned (applicable to other regions)

- is possible to amaglamate different ecosystem information into one place
- not an off-the-side-of-your-desk project
- is being used by DFO scientists (and others)
- infrastructure of the package could be utilised in other regions if interested
  - e.g. my monthly download of buoy data is for all Canadian waters, the data processing, storing, and plotting functions could be adopted
  - similarly for other types of data/information

Summary

- https://github.com/pbs-assess/pacea
- installation: `remotes::install_github("pbs-assess/pacea")`

We wrangle the data so you don't have to

```
citation("pacea")

Edwards AM, Tai TC, Watson J, Peña MA, Hilborn A, Hannah CG, Rooper CN
(2023). "pacea: An R package of Pacific ecosystem information to help facilitate an ecosystem approach to fisheries management."
  https://github.com/pbs-assess/pacea.
```