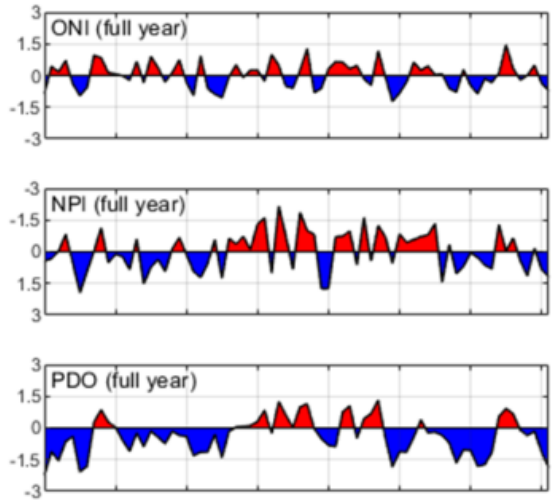
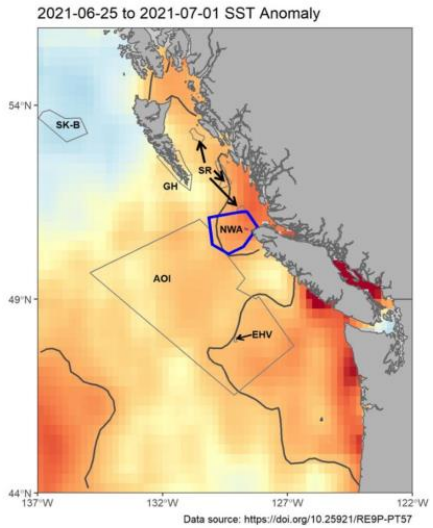
Introducing pacea: an R package to amalgamate Pacific data to help operationalise an ecosystem approach to fisheries management

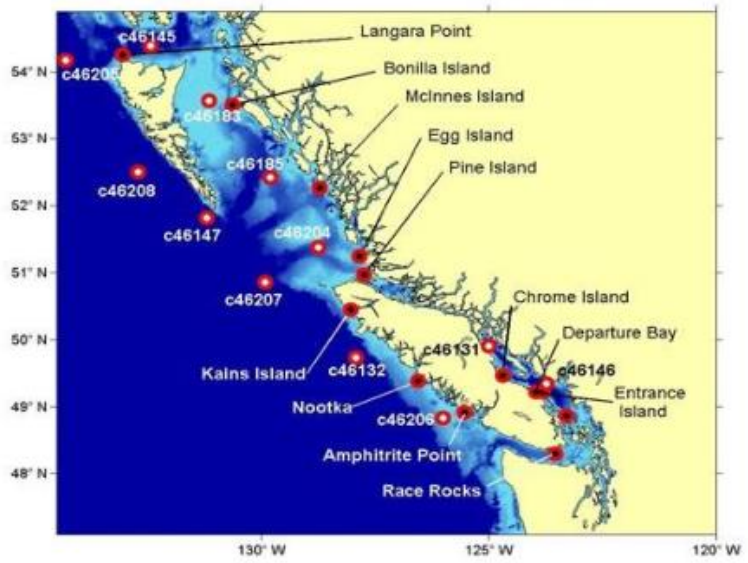
Andrew Edwards & Travis Tai

Pacific Biological Station, Fisheries and Oceans Canada, Nanaimo, BC.



pacea





Government of Canada / Gouvernement du Canada

Canada

EBM MSP talk
Friday 13th October 2023

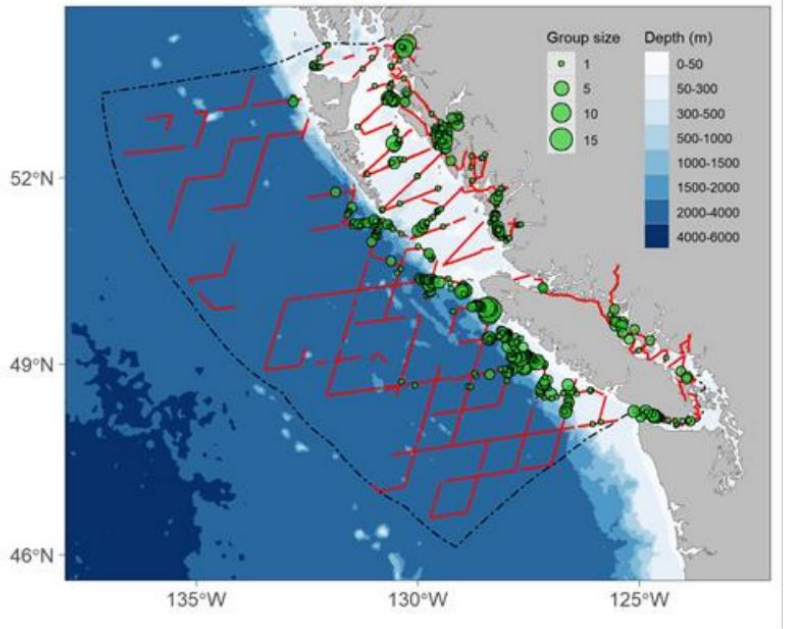
Motivation

- 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 et al. (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

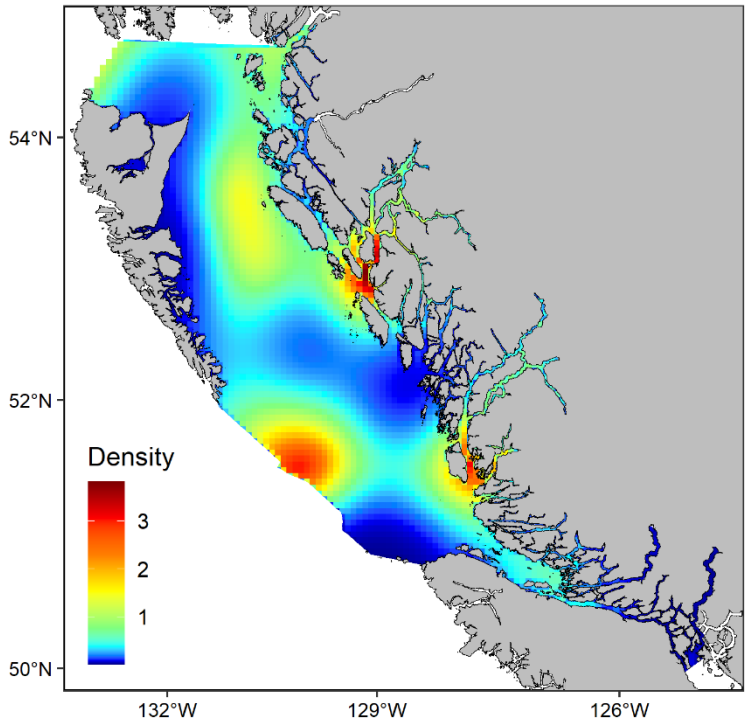
Motivation (based on a true story)

Survey sightings of Humpback Whales



Modelling

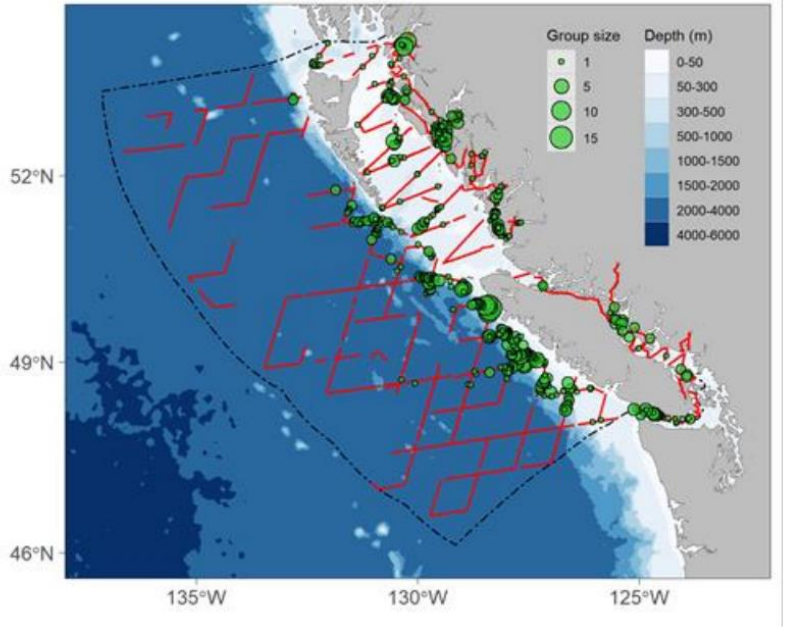
Estimated densities



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

Motivation (based on a true story)

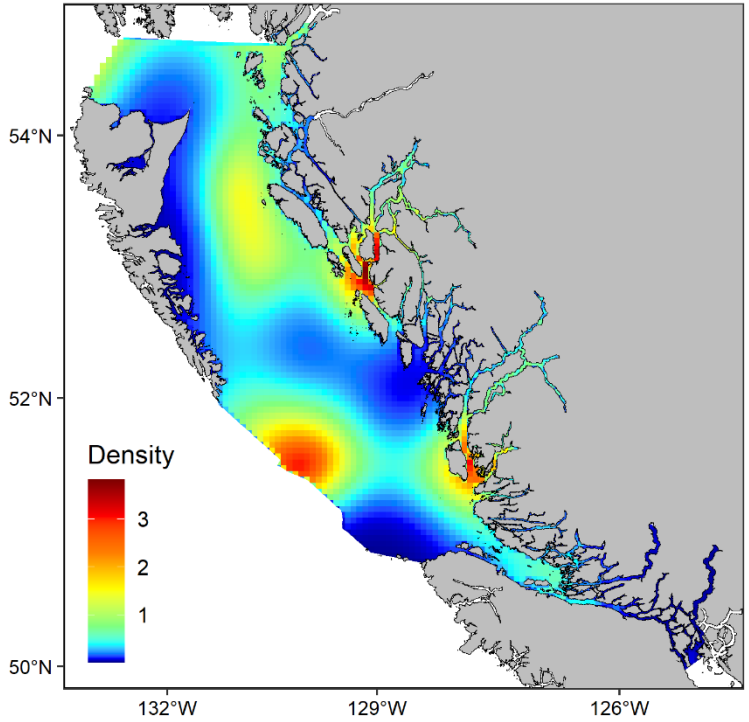
Survey sightings of Humpback Whales



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	WMS	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 some data wrangling to be usable – usually takes way, way, way longer than anticipated.

So SST analysis did not happen.

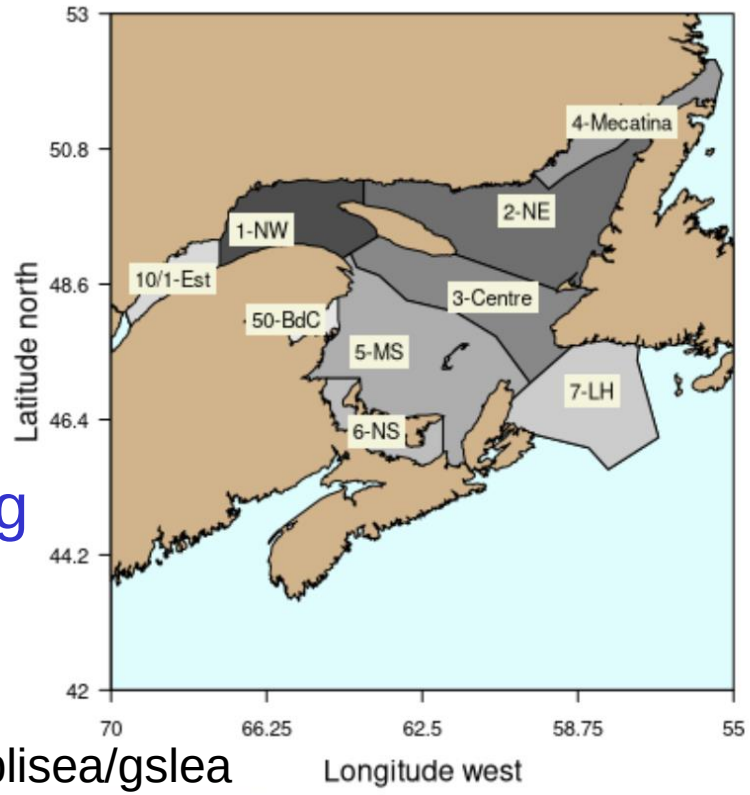
Motivation

- Similarly, ROMS (Regional Ocean Modelling System) outputs are available from Angelica Peña (a DFO oceanography).

- But still requires extensive coding to convert results from netCDF files into R.

Motivation

- Similarly, ROMS (Regional Ocean Modelling System) outputs are available from Angelica Peña.
- But still requires extensive coding to convert results from netCDF files into R.
- “Open Data” is not enough.
- Hence, the pacea R package, motivated by the GSLea package for the Gulf of St. Lawrence.
- Primary audience is Pacific DFO stock assessment scientists, but usable by anyone (need 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

Spatial data

Spatial data will be stored on a master grid (likely 2 km x 2km), conceptually similar to:

Currently working on finalising the master grid size.

PACea fundamentals

It's an R package

Contains temporal, spatial, and spatiotemporal datasets

All data fully documented, including citations and sources, and vetted by experts

Fancy spatial stuff done behind the scenes – users can export as simple data frames or .csv files

Traceable and updatable – each dataset to have reproducible code showing how it was made

Fully open source and downloadable from GitHub

Vignettes showing example applications and plotting

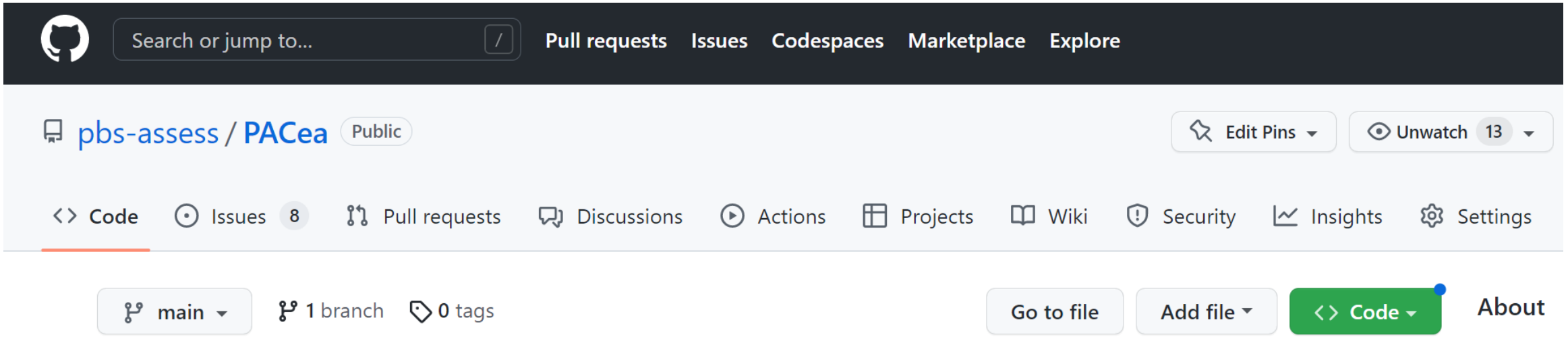
Continually expand with more datasets

Example datasets initially being incorporated

- ROMS output from Angelica Peña:
  - temperature at the sea surface and at various depths
  - oxygen at various depths
  - salinity at various depths
  - primary production
  - chlorophyll
  - pH at different depths and depth of aragonite saturation
- sea-surface temperature from satellite measurements (one of the ERDDAP datasets)
- lighthouse temperature time series
- oceanographic indices such as Pacific Decadal Oscillation, Oceanic Niño Index, Southern Oscillation Index

Monthly resolution where possible.

Openly available at github.com/pbs-assess/PACea



- Not yet operational, but once it is we will ensure it remains usable even as we expand it.
- Dealing with spatial data is tricky (many options and R packages), but users will be able to use simple data frames of lon, lat, and (say) sea-surface temperature.

What PACea is *not*

- Not going to be the sole repository for detailed raw data (not replacing existing databases).

- Not replacing Open Data, but utilsing it and making it easier to incorporate into analyses.

“Organise your data you'll find those connections” – Mark Leblanc (CHS), emphasising benefits of storeable and shareable data (Oct 2022 All Staff).

Acknowledgments

Thanks to Angelica Peña, Kelsey Flynn, Chris Rooper, Lindsay Davidson, Jessica Nephin, Brianna Wright, Charles Hannah, Patrick Thompson, and others for earlier feedback at the planning stage. CSRF for funding.