Data Lab #3: The Blob

In this data lab, you will combine temperature data from multiple sources to study a recurring marine heatwave in the North Pacific Ocean, known as The Blob.

We have discussed some introductory background on marine heatwaves and The Blob in class, but for additional information (which will ultimately be helpful for your lab writeup), here are some sources that I recommend as a starting point:

- Excellent compilation of resources on marine heatwaves from an international working group of scientists studying this subject: http://www.marineheatwaves.org/
- Article from NOAA last fall reporting on the effects of The Blob and potential signs of its return this year: https://www.fisheries.noaa.gov/feature-story/new-marine-heatwave-emerges-west-coast-resembles-blob
- Article from NASA focused on how satellite sea surface temperature data have helped us better understand The Blob: https://earthdata.nasa.gov/learn/sensing-our-planet/blob

Part 1: Acquiring and extracting data

Key skill goals for Part 1 of this lab are for you to practice accessing and working with raw – and often messy! – data, including both extracting and exploring netCDF data files in MATLAB and learning how to find and download data from online data repositories.

In this lab, we will be interpreting ocean temperature data from three different sources:

1) Data collected from a sensor deployed by the Ocean Observatories Initiative
2) Satellite observations of sea surface temperature compiled by the National Oceanic and Atmospheric Administration
3) A gridded interpolation of a compilation of many data points, known as the World Ocean Atlas (similar to the gridded data in the LDEO ocean pCO_2 dataset we used in Data Lab #2)

In Part 1, you will acquire, extract, and begin working with the data from the first two sources. In Part 2, you will combine and compare all three of these datasets to determine what we can learn from them about The Blob.

1) Data from the Ocean Observatories Initiative (OOI)

The first dataset you will use for this lab comes from a Conductivity-Temperature-Depth (CTD) sensor on Subsurface Flanking Mooring B at the OOI Station Papa array. You can find lots of information about the OOI program, and the specific array, mooring, and sensor that these data come from at: https://oceanobservatories.org/array/global-station-papa/

Getting the data:

- To save time and avoid all of you having to separately download too many files, I have pre-downloaded the data files you need and provided them for you at: https://tinyurl.com/BlobLabOOIdata.

- You will need to download these files and have them in your MATLAB path to work with them – but don't put them inside your GitHub repository because they take up too much space!
- If you're curious about where I got the data from for you, they were downloaded from the OOI Data Portal at https://ooinet.oceanobservatories.org/

Getting started with these data in MATLAB:

- Starter code with step-by-step directions for what you will do with these data is in GitHub classroom at: https://classroom.github.com/a/PDx9kfF7
- Though I encourage you to pick a new partner to collaborate with for this lab, you will each have your own GitHub repository and write the code yourself.
- Note that this is the first time in this class that you are working with original raw data! That means that the data will be messy and imperfect for the task at hand. In Part 1, you will read in, explore, and complete preliminary processing of this original raw data to prepare to calculate temperature anomalies in Part 2.

2) Satellite observations of sea surface temperature

One of the learning goals for this class is for you to be able to find and use publicly-available data sources. One great resource are servers hosted by reliable sources that compile many different types of data. There are many different servers where you can find this type of earth and environmental science data, but one that is particularly valuable is the ERDDAP server that provides access to National Oceanic and Atmospheric Administration (NOAA) data. To get the second dataset for this lab, you will download it yourself from this ERDDAP server!

First, though, you will start by learning about ERDDAP and how to access data through it – a skill that I anticipate will be useful for many of you for your final projects.

Getting started with ERDDAP:

- Take a look at the introduction page about ERDDAP at: https://upwell.pfeg.noaa.gov/erddap/index.html (note that a lot of this is very technical so I don't expect most of it to make sense – this is also an exercise in consuming publicly-available data, which includes sorting through a lot of jargon and often-hard-to-follow detail)
- Read through Chapters 1 and 2 (each very short!) of NOAA's ERDDAP tutorial at: https://coastwatch.pfeg.noaa.gov/projects/erddap/introduction.html. If you are interested in using ERDDAP for your project, you may want to complete more of the tutorial.
- Follow the step-by-step directions in 5.1 and 5.2 of the tutorial (https://coastwatch.pfeg.noaa.gov/projects/erddap/hovmoller.html#visualize-the-blob-with-sst) which will specifically introduce how to access satellite sea surface temperature and temperature anomaly to interpret in the context of The Blob.

Downloading your own data file from ERRDAP:

- You will now download a subset of the sea surface temperature anomaly data (jplMURSST41anommday) that you accessed and plotted in tutorial section 5.2 above.

- Instead of clicking on "graph" in the "Make a Graph" column, now instead click "data" in the "Grid DAP Data" column.
- You have two goals in downloading data:
    - Temporal coverage: monthly data from January 2013 through February 2020
    - Spatial coverage: a map of the northeastern Pacific, centered around Ocean Station Papa, covering at least 30 degrees each of latitude and longitude
- **IMPORTANT!** If you download the data at its original resolution (1/100 of a degree) the size of the data will be HUGE (> 1 GB). You will need to subset the data.
    - In the "stride" parameter, reduce the spatial resolution of the data by setting a stride value >1 (i.e. a value of 100 would provide data at 1 degree resolution). I'd recommend starting with 1 degree resolution data to check that you're not accidentally downloading too huge a file, though if your download speed and laptop can handle it, I'd recommend using ¼ degree resolution for the lab analysis itself (which is a 17 MB file).
- Once you have your space and time bounds selected, check that you want the "sstAnom" variable and pick ".nc – Download a NetCDF-3 binary file with COARDS/CF/ACDD metadata" from the File type dropdown menu. Then click Submit to get your data!

Opening and checking your data in MATLAB:

- Start a new MATLAB script in the Blob Data Lab Part 1 GitHub repository to open and explore this new data file
- I will give you more explicit instructions for Part 2, but your goal here – following the model for how you extracted data from the OOI netCDF files – is to:
    1) Check that the data has the time, latitude, longitude, and sstAnom data that you had intended to download
    2) Make a map of the sstAnom from February 2020 to check that everything looks reasonable.
        - Hint: check the data quickly first by using the function "imagesc". Then when you're ready to make a more polished map, look back at the code from Data Lab #2 using the Mapping Toolbox as a model – you can specify plotting only part of the globe setting the map bounds – i.e. "worldmap([20 60], [-170 100])