National Oceanography Centre DSG Challenger: Eddy Detection

Overview

Mesoscale eddies are ubiquitous in the global ocean. They are large rotating vortices, typically with spatial scales of 100 km and timescales of 30 days and play an essential role in transferring heat, energy, and material around the ocean. On a global scale, the movement of eddies accounts for approximately 90% of the ocean's kinetic energy, and they are known to affect ocean dynamics, weather conditions and commercial activities. Eddies are also explicitly linked to climate change, playing a crucial role in the deep storage of carbon and heat, whilst long-term shifts in poleward eddy activity are indicative of a warming climate. As a result, being able to track and detect eddies is crucial for physical oceanographers in not only understanding eddies but also in accurately modelling the changing oceans and climate.

Being able to detect and track the movement of eddies is also vital in gathering data on eddies. Oceanographers who wish to study the depth profiles of eddies collect data by crossing through them with aquatic robots such as gliders or autonomous underwater vehicles, measuring temperature-salinity and other parameters which describe their behaviour and transport properties. For other studies, knowing the locations of eddies can help pilots avoid getting stuck in them or by using them to their advantage by using the currents to increase the endurance of glider missions.

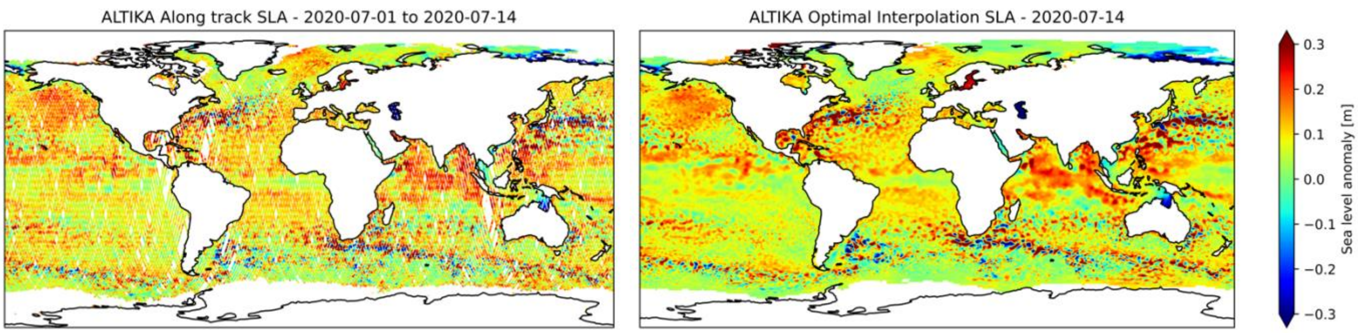

Figure 1: a) Global along-track measurements of sea-surface height from the Altika satellite over a 2 week period. Note the orbits of the satellites and sparisty of observations. b) 1/4° resoltuion gridded sea surface height map from AVISO/CMEMS.. The map is derived by optimally interpolated the along-track measurements onto a regular lat-lon grid with a 1 day temporal resolution. You can clearly see the eddy currents which are denoted by areas of high/low sea level anomaly (current sea-surface height minus a 5 month moving average)

Detecting eddies is however non-trivial. According to the direction in which they rotate eddies can cause significant upwelling or downwelling which leaves strong signatures in sea surface height (SSH) which are detectable using gridded satellite altimetry products. Such gridded products involve interpolating sparse sea surface height observations from a constellation of orbiting satellites onto a regular latitude-longitude grid using *optimal interpolation (OI)*. OI is widely used method of data assimilation that provides an estimate for the current sea surface height via a weighted least squares fit to observations. The weights in OI are the inverse of the error covariance matrices for the observations and the background field.

With access to a gridded SSH map, most studies then rely on an automatic detection algorithm that identify and track the eddies. These methods can be typically divided into parameter-based models and more recently machine learning based approaches. Parameter or model-based methods typically look for maxima and minima in sea-level anomaly maps, defined as the difference between the current SSH minus a 5-month moving average, before forming a closed contour around the extremum. Machine learning methods on the other hand often adopt approaches from computer vision such as image segmentation or object detection algorithms. Currently, the most used approach is to use software called *py-eddy tracker (PET)* which is a model-based approach that iteratively builds eddies under a set of rule-based conditions.

Challenge

However, there is an open question as to how well current eddy detection methods work. Firstly, global gridded SSH maps which are used to detect eddies are only provided at $1/4°$ resolution (~ 30 km grid spacing). At these scales unresolved eddies have been shown to introduce aliasing effects that distort the characterisation of larger eddies, resulting in underestimation of smaller eddies and over estimation of larger ones. Indeed, Amores 2018, show that satellite-like products capture roughly 6% of actual eddies in the North Atlantic. Consequently, these flaws in altimetry-based products have significant impact in downstream scientific research. There exist many studies which have used altimeter data to make inference on eddy properties, relate eddies to other ocean variables, and to estimate surface eddy kinetic energy. In all cases therefore, the results must therefore be treated with caution given most of the eddy field is missed and surviving parts are likely to be affected by noise and aliasing.

Further detecting eddies from SSH maps is also tricky. With very few labelled examples, machine learning approaches have often been hindered by a lack of high-quality data, tending to use extensive data-augmentation or by training on the outputs of PET. Training on the outputs of PET however is problematic as your machine learning algorithm will try to learn the biases and shortcoming of PET. Model-based methods are also prone to errors, requiring one to define a strict set of criterium which must be satisfied in order to classify an eddy. This can lead to eddies being erroneously split into separate vortices and model-based detection algorithms being sensitive to observation noise. Further, eddies are also prone to disappearing before reappearing again, making it hard to ensure that the temporal consistency of eddy paths.

Consequently, the aim of this challenge is to improve using machine learning how we can detect and track eddies. We split this problem into 2 parts as described above:

a. SSH Mapping - *How can we accurately map SSH onto a high-resolution grid given a set of sparse satellite observations using machine learning?*
   I. From sparse satellite observations are we able to accurately map to a gridded resolution of greater than $1/4°$?
   II. With better mapping are we able to resolve mesoscale eddies more effectively?
   III. Is machine learning an effective way to resolve issues with aliasing and distortion found when applying optimal interpolation?

b. **Eddy Identification -** *How can we effectively detect eddies given a SSH map using machine learning?*
   I. Can we improve the robustness of eddy detection algorithms to noise and artifacts in SSH maps?
   II. Can we reduce our reliance on labelled data by using data augmentation or techniques from unsupervised learning?
   III. Could we quantify the uncertainty in our predictions when detecting and tracking eddies?

For the challenge teams should pick **either challenge (a) or challenge (b) but not both.**

Both challenges involve improving a current part of the workflow in detecting and tracking eddies and both challenges could benefit from suitable and tailored application of machine learning. Both challenges will also be tested in how well they can resolve mesoscale eddies from sparse along track observations:

a. **SSH mapping** – method will be tested by passing your gridded SSH map to PET which we will then be used to detect eddies.

b. **Eddy Identification** – method will be tested by using $1/4°$ degree optimally interpolated maps and then applying your eddy detection algorithm.

Data

Arguably, the biggest hindrance to improved SSH mapping and eddy detection is the lack of an accurate high-resolution ground truth field. Recently, there have been significant advances in large Ocean General Circulation Models (OGCMs) which resolve currents at very high resolutions. As a result, these models have been used increasingly as a proxy for high resolution ground truth data such as SSH.

In this project we will use the NEMO-eNATL60 North Atlantic Ocean simulation, as our high-resolution ground truth for the SSH field. The model output resolution of $1/60°$ and is designed to model as accurately as possible sea surface signatures down to 15 km. By using the model our assumption is that the dynamics our representative of the real-world ocean however, the state is not. Consequently, we will also provide SSH observations corresponding to the Altika, H2B, Jason 3, Sentinel 3a and Sentinel 3b satellites by interpolating the high-resolution model onto their real-life paths.

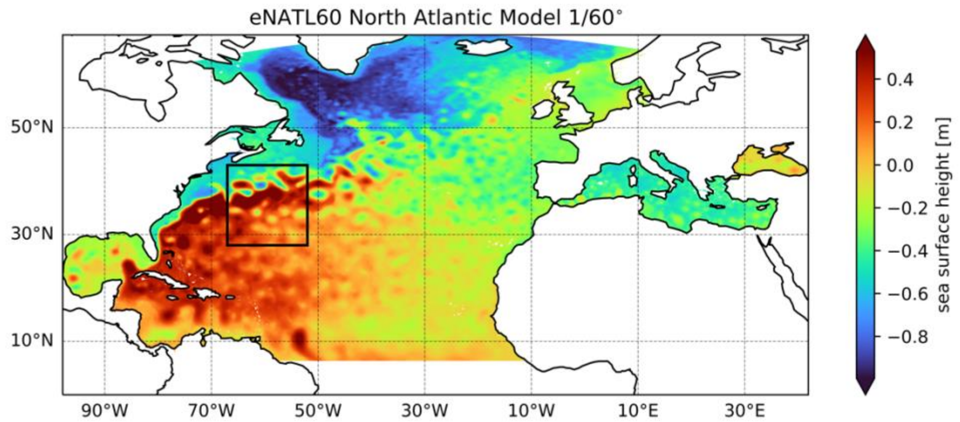

Figure 2: SSH from the NEMO-eNATL60 ocean model at $1/60°$ resolution. The black box denotes the restricted region which we will use for this challenge.

In this project we will also only consider a $1/15° \times 1/15°$ section of the North Atlantic near the Gulf Stream. This area is picked as is an active area of eddy activity. A summary of the provided data is given below:

1. **High-Resolution Ground truth** – Derived from eNATL60 simulation and regridded onto a regular latitude-longitude grid at a $1/12°$ resolution. The data has a 1-hour temporal resolution spanning 2020-09-01 to 2021-06-30.

2. **Simulated Along Track Measurements** – Derived by linearly interpolating the model outputs onto the paths from real life satellite tracks from the Altika, Jason3, H2B, Sentinel 3a and Sentinel 3b satellites, spanning 2020-09-01 to 2021-06-30.

3. **Ground Truth Eddy dataset** – Derived use PET applied to the high-resolution ground truth dataset.

Goals

Depending on the backgrounds and interests of the group, the group can choose to explore either challenge. The goals of both challenges are to improve eddy detection using novel applications of machine learning. Some potential outcomes or suggestions for both challenges are listed below:

a. SSH Mapping
   i. Apply methods from computer vision to learn SSH field
   ii. Use methods from in-painting to handle sparse observations
   iii. Use probabilistic methods to quantify uncertainty in regions where we have less data
   iv. Use interpretable methods such as Gaussian processes to capture both spatial and temporal non-stationarities in the data
   v. Develop an algorithm that is scalable such that it can be applied to global domains and not just restricted ones.

b. Eddy Identification
   i. Apply methods from image classification, object detection or image segmentation to detect eddies from SSH maps
   ii. Use unsupervised or semi-supervised techniques to reduce reliance on labelled data
   iii. Generate more training data using generative models or data augmentation.
   iv. Use probabilistic methods to quantify uncertainty on whether an eddy exists or not
   v. Use temporal data or auto-regressive models to help predict locations of eddies

What we're looking for

For this challenge we welcome participants from all backgrounds who are interested in making a tangible difference to current oceanographic research. We predict that prior experience in any one of the following areas might be particularly helpful in solving the presented challenges:

- Image classification/object detection/image segmentation
- Spatiotemporal modelling
- Probabilistic machine learning
- Continuous time machine learning
- Handling earth science data using netcdfs
- Climate or oceanography research
- Unsupervised learning