Mapping and Visualizing the Ocean Using Scattered Glider Observations: A Toolset to Investigate Submesoscale Ventilation

Dhruv Balwada

Since the industrial revolution 25% of the human-created carbon and 90% of the excess heat in the earth system has been sequestered into the deep ocean. This carbon and heat is transported from the surface ocean into the deep ocean in a process referred to as ocean ventilation. Most of this ventilation is a result of ocean flows lasting less than ~10 days and shorter than ~20km (submesoscale flows). However, our knowledge of the submesoscale flows, particularly their sub-surface structure, is still in its infancy due to the numerous challenges associated with observing the ocean's interior at these scales. Understanding these submesoscale flows is particularly important since these scales are too small to be resolved by our current climate models; the fidelity of climate predictions hinge on accurate parameterizations of their effects.

Ocean gliders, which are guided autonomous observational platforms, navigate the ocean by profiling the water column vertically to depths of ~1km and using this motion to propel themselves horizontally by ~5km every 6 hours. Observations are collected using a variety of sensors along these profiles, providing a high resolution (millions of data points), but scattered, spatio-temporal measurement of the state of the ocean. In principle these glider observations can be used to gain insight into the nature of submesoscale flows by identifying features in the data at these scales and doing an in-depth analysis of the properties of the identified features. This might also include looking for connections between the identified features at depth and observations of the surface properties made using satellites. However, to gain insights from the glider data a number of technical challenges need to be addressed.

The first challenge: How do we map the scattered spatio-temporal observations to a regular grid and how do we estimate the uncertainty in the mapped estimates? Having a mapped estimate is an essential first step in further analysis, which might involve identifying coherent structures, estimating gradients, or gaining insight in to the multi-scale structures via spectral analysis. Most recent glider-based analyses have used very simple mapping techniques, such as linear interpolation or kriging with fixed parameters, which are easy to implement but do not necessarily respect or learn from the observed structure in the data. A novel way to address the mapping challenge would be to use machine learning, eg Gaussian Process Regression (GPR), which can learn the optimal mapping parameters from the available data and also produce robust error bars on the mapped estimates.

These mapped observations are helpful to identify interesting features in the data, which might be indicative of individual ventilation events resulting from submesoscale flows and can be studied further. However, this process is very tedious, since one usually needs to simultaneously investigate many different variables, potentially from different sources, and also use a variety of plotting parameters to find what works best; even with a very systematic manual approach one can very easily miss important features. This brings us to *the second challenge*: build an interactive and dynamic visualization dashboard, which significantly reduces the burden of manual plotting and helps insights to emerge more fluidly. One way to address this would be to use the Holoviz tools in python (https://holoviz.org/), which allow for complex and interconnected dashboards to be built. Also, the implementation should be done in a way that other oceanographers, with some basic knowledge of python, can easily customize the dashboards to their specific needs.

My proposed project for the eScience Winter Incubator has two deliverables:

1. An explorable implementation of a machine learning based technique for mapping scattered glider data, which learns parameters from the data.
2. A dashboard for interactively visualizing glider data along with other complementary datasets (eg satellite data or historical ship data). The implementation should be modifiable relatively easily by oceanographers with intermediate level python skills.

Both these deliverables would benefit from a partnership with a data scientist who possesses in-depth knowledge of modern data science tools and practices, as they require the development or implementation of software that can handle large data sets and is optimized to rapidly produce results. Although I have identified a few potential avenues that can be taken to achieve these deliverables (gaussian process regression and holoviz), I am also open to trying and exploring other paths as well.

This proposed project will facilitate the investigation of glider observations, using modern data science techniques, to understand the dynamics and effects of submesoscale flows in ventilating the ocean. The immediate tangible results, a more robust mapping of the observations and an interactive visualization platform, will benefit the broader community of oceanographers using gliders to investigate other research questions. In the future, these insights will help guide the design of the next generation of ocean parameterizations in climate models.