Biogeochemical modelling and data assimilation: status in Australia and internationally

Richard Matear, Andrew Lenton, Matt Chamberlain, Mathieu Mongin, Mark Baird
CSIRO Marine and Atmospheric Research, Hobart, Australia

1. Introduction

Marine biogeochemical (BGC) modelling is a key approach to helping us understand the biochemical processes responsible for the transfer of nutrients and carbon between inorganic and organic pools (Matear and Jones, 2010). Quantifying these biochemical processes is essential to understanding carbon cycling in the ocean, the air-sea exchange of carbon, the impact of climate variability and change on marine ecosystems, and the link between ocean physics and ocean biology.

Data assimilation represents a new and exciting tool to advance ocean biogeochemical studies. The coupling of physical and biogeochemical data assimilation is a natural evolution of the Global Ocean Data Assimilation Experiment (GODAE). Data assimilation fuses models with a diverse set of observations to provide a more consistent view of the physical and biological state of the ocean. Extending the GODAE effort to include BGC data assimilation presents an exciting opportunity to expand the researchers interested in using ocean data assimilation products and Brasseur et al. (2009) provides a summary of BGC applications of data assimilation. However, delivering these new BGC data assimilation products is not a trivial task. In this presentation, we discuss the challenges of utilizing the existing physical ocean data assimilation systems to include BGC data assimilation. To help set the context of this discussion, we will first briefly discuss previous BGC data assimilation effort. Then, we will discuss both international and Australian BGC data assimilation activities.

2. Biogeochemical Data Assimilation

The application of data assimilation separates into two types of problems: 1) parameter estimation and 2) state estimation. The two approaches reflected different philosophies on how to fuse the BGC model with the observations with both delivering useful but different information. In both approaches, data is used to constrain the evolution of the state variables but with parameter estimation it is the model parameters that are modified to fit the data constraints while in the state estimation the state variables are modified to fit the observations. The state estimation approach generally forms the foundation of GODAE physical data assimilation with such effort delivering either ocean forecast or reanalysis products of the time evolving 3-D ocean physical state. However, BGC data assimilation studies have tended to focus on data assimilation for parameter estimation. Although a complete description of the two approaches is beyond the scope of this presentation, the following is a brief summary of the application of these two data assimilation approaches.

2.1 Parameter Estimation

An obvious feature BGC models are the large number of model parameters that must be specified to simulated the BGC cycles. The setting of these parameters can be partially accomplished from observations but for many parameters no direct observational estimate is available. Further, even for the more observable parameters like the phytoplankton growth parameters much uncertainty still exists in setting the parameter values because either the parameter values change with time or the values determined from the individual phytoplankton species may not be applicable to the entire ecosystem. The inability to directly specify all the model parameters forces one to determine the parameters values by tuning the model to reproduce the observations. This is a tedious and time consuming approach. The attraction of data assimilation is that it provides a means to generate a set of
model parameters that reflects the observations, determines the values of the poorly known parameters, and provides insight into which parameters are constrained by the model (Kidston et al., 2011; Schartau and Oschlies, 2003).

There is now a long list of studies which have used data assimilation methods to estimate model parameters of ecosystem models (see Gregg et al. (2009) for a list of these studies). There is even a webpage setup to explore parameter estimation of a suite of different BGC models at a number of different sites (http://www.ccpo.odu.edu/ marjy/Testbed/Workshop1.html ).

The uncertainty in the biological model is not just in the model parameters but extends to the choice of equations used to describe the biological system (Franks, 2009). Both parameter and model formulation uncertainty introduce large model error in BGC modelling, which lends itself to parameter estimation studies which explore the parameter space, model complexity and model formulation (e.g. Matear (1995)). The ability to address all three of these issues demonstrates the value in the parameter estimation approach for BGC data assimilation. In addition, parameter estimation can provide insight into what observations are critical to building and constraining more realistic models, and identifying the critical model parameters required to reproduce the observations (Kidston et al., 2011). The latter result provides a convenient way to identify a subset of critical model parameters, which capture the key observed dynamics of the biological system, and then explore how these parameters affect the dynamics of the system (e.g. Friedrichs et al. (2006)).

One important aspect to note with the parameter estimation approach is that unrealistic parameter values may be estimated because important processes are excluded from the model formulation (these are call structural errors in the model). Therefore, the estimated model parameter values must be ecologically assessed and deemed plausible otherwise the formulated model has structural errors.

### 2.2 Ocean State Estimation

The attraction of state estimation data assimilation is that it provides a way to incorporate both physical and biological observations into the numerical models to obtain the evolution of BGC fields that are dynamically consistent with the observations and provide a tool to extend the observations in both space and time (Lee et al., 2009). The application of state estimation is to overcome limitations in the model by correcting the ocean state to produce a more realistic evolution of the ocean state (Natvik and Evensen, 2003a). The approach provides a way of limiting the impact of model errors (parameter, formulation, initialization and forcing) to better hindcast and forecast the ocean state.

The study of Gregg (2008) provides a nice review of sequential BGC data assimilation studies which I briefly summarize here. Not surprising the focus of these studies is on assimilating ocean colour surface Chlorophyll a concentrations into their BGC models since this data product provides the best spatial and temporal data of the ocean biological system.

The first example of sequential data assimilation directly inserted CZCS chlorophyll into a 3-dimensional model of the southeast US coast (Ishizaka, 1990). They produced immediate improvements in their chlorophyll simulation but these improvements did not last more than a couple of days before the model simulation diverge from the observations. The divergence of the model simulation from the observations reflected a bias in the biological model to overestimate Chlorophyll a. Correcting such a bias would be crucial to obtaining better and longer lasting state estimation results. More recently, the Ensemble Kalman Filter
(EnKF) has been used to assimilate SeaWiFS ocean color Chlorophyll a data into a 3D North Atlantic model (Natvik and Evensen, 2003a, 2003b). They showed the EnKF updated ocean state was consistent with both the observed phytoplankton and nitrate concentrations. However, this study did not show any comparison to unassimilated data to enable a quantitative assessment of the impact BGC state estimation (Gregg et al., 2009). Recently, Hemmings et al. (2008) presented a nitrogen balance scheme with the aim of assimilating ocean color Chlorophyll a to improve estimates of the seawater pCO2. They used 1D simulations at two sites in the North Atlantic (30N and 50N) to assess the performance of their scheme. The scheme exploits the covariance between Chlorophyll a and the other biological state variables at a fraction of the computational cost of the multivariate EnKF scheme. To do this Hemmings et al. (2008) use 1D model simulations with varying model parameters to extract the relationship between simulated Chlorophyll a and the other biological variables. This information was then used to project the assimilated Chlorophyll a data on to all biological state variables. They show the nitrogen balancing approach improves the ocean pCO2 simulation over the case where only phytoplankton and DIC are updated.

At present, the application of state estimation has been confined to utilizing remotely sensed ocean colour Chlorophyll a. To extend this information from the surface into the ocean interior will require coupled biological -physical models. Further, many of the colour images are corrupted by clouds and filling these data gaps is another obvious outcome of state estimation. Finally, it quite possible that the fields that we are most interested in are not sufficiently observed to provide the spatial and temporal coverage desired (e.g. pCO2). For these cases, state estimation data assimilation provides tool to exploit the existing observations to generated the fields of interest.

Data Assimilation Activities
To help articulate existing data assimilation activities both internationally and within Australia we draw from the international of the GODAE Ocean View program, which has set-up a new Task Team called Marine Ecosystem and Prediction ( MEP). CSIRO is involved in the MEP- Task Team. The MEP Task Team seeks the integration of new models and assimilation components into operational systems for ocean biogeochemistry and marine ecosystem monitoring, which aims to bridge the gap between the current GODAE OceanView capabilities and new applications in areas such as fisheries management, marine pollution, water quality and carbon cycle monitoring.

3.1 GODAE (International) Activities
Exploiting Existing GODAE Ocean View Products
Operational Oceanography (OO) products, both hindcasts or reanalyses of the ocean state provide the best representation of the ocean state hence such products should be used by BGC models. However, there is a lack of documented studies explicitly discussing the practical advantages and weaknesses of using constrained versus unconstrained physical forcing to drive biogeochemical models. The key activity here is to provide a more accurate picture on how biogeochemical modellers can benefit from the data products and insight already being generated by GODAE OceanView.

This activity will also investigate how GODAE OceanView products can be improved for BGC simulations. For instance, it is widely recognized that the GODAE systems need to improve their representation of physical variables, such as the vertical fluxes in the upper ocean, which are critically important to biological processes. This activity will require direct inputs and efforts from experts in ocean circulation modelling and data assimilation involved in the development of operational systems to interface with BGC modellers to produce improvements that benefit both physical and BGC simulations.
b) Biogeochemical model development and observing systems

The first aim of this activity will focus on 2 aspects: (i) the downscaling from global to regional systems, by provision of biogeochemical boundary conditions and the (ii) further development of 2-way biogeochemical coupling in models to assess bio-physical feedbacks and quantify the impact of biological activity in, for example, heat fluxes.

The second aim is the development of multi-purpose observing systems. Key to this activity will be to identify the essential physical and biogeochemical observations required to constrain the coupled physical and BGC models and to formulate relevant recommendations to further develop the global ocean observing system. The new observations should target of several key biogeochemical applications, such as ocean CO2 flux monitoring, Harmful Algal Blooms and fisheries management.

**Australia Activities**

The Australian activities are closely integrated with the GODAE MEP activities, and the jointly funded CSIRO/BOM/RAN Bluelink 3 project with additional effort focused on the coastal environment.

The Bluelink 3 project will provide the key project for delivering results to GODAE Ocean View activities. With Bluelink 3, several new initiatives were planned. The ocean model is now eddy-resolving in its entire domain (1/10 degree resolution between 75S to 75N). A simple Biogeochemical cycle (BGC) module with phytoplankton, zooplankton, nitrate, detritus, oxygen, and carbon has been added to the ocean model to enable the simulation of phytoplankton and carbon. A re-analysis product with BGC fields will be produced along with an assessment of how physical data assimilation impacts the BGC (GODAE activity a).

The Bluelink 3 project will also deliver regional simulations with the global model providing the boundary conditions for the open boundary of the regional (e.g. Great Barrier Reef) and local (e.g. Heron Island) simulations. Both the global and regional models will deliver both forward running and data assimilation products with BGC cycling in the water column and in the sediments. The Bluelink 3 effort intends to exploit the data streams generated by the Integrated Marine Observing System to assimilate into the model and to assess our model simulations. These data streams will include, coastal BGC reference sites, argo drifters, ocean colour products, gliders and moorings.

In addition to the Bluelink project, data assimilation is also occurring in ACCESS-o (the coupled ocean and sea-ice model that Australia is using for its AR5 climate change simulations). With ACCESS-o, a multi-decadal re-analysis product will be generated. Several test re-analysis products have been produced and we expect within the next year to complete a re-analysis simulation with BGC (using the same BGC model that is used in Bluelink 3). The BGC data assimilation will assimilate remote sensed ocean colour Chlorophyll a to constrain the BGC cycle parameters.

The funding of the Wealth from Ocean Carbon Cluster has provided a third data assimilation effort in Australia. The effort requires extending the coastal BGC model to include carbon and the key processes and time-scales required to simulate the carbon cycle in the Australia coastal waters. This new effort will be challenged to exploit limited BGC understanding and observations to develop a realistic representation and simulation of carbon cycling in coastal waters. For this application, BGC data assimilation will be used to assist in parameterising the BGC model, to pursue state estimation of BGC fields and to direct the observing strategy of the both the Carbon Cluster and the IMOS.
Conclusion
The field of BGC data assimilation is a relatively new but there are now many examples where the approach has been applied to both parameter estimation and state estimation problems. Data assimilation with BGC models provides a framework to extract information from BGC observations and refine prognostic models of carbon and nutrient cycling in the ocean. The existing GODAE data assimilation systems are an obvious avenue for expanding data assimilation to include BGC. At present the modifying the GODAE system to include BGC is an important focus of CSIRO data assimilation effort but through the Carbon Cluster the BGC data assimilation effort in the coastal environment will expand.

At present, remotely sensed ocean colour Chlorophyll a is the key data stream to assimilate but we need to evolve the BGC models to actually simulate the observed quantity (e.g. for ocean colour Chlorophyll that is water leaving radiances). Further, better use of other data streams needs to be explored and data assimilation can help identify the optimal observations required to constrain the BGC model for a given problem.

Data assimilation has the potential to improve BGC models, deliver better estimates of BGC ocean state and direct observational strategies. All 3 tasks are important but do them requires continued access to data streams like remotely sensed ocean colour and the allocation of sufficient human and computation resources to tackle these demanding problems.

References


