Towards a Data Assimilative Physical-Biogeochemical Ocean Model

Paul Mattern[1,*], Jill Falkenberg[2], Michael Dowd[2], Katja Fennel[1]

[*] Paul.Mattern@dal.ca
[1] Department of Oceanography, Dalhousie University
[2] Department of Mathematics and Statistics, Dalhousie University

Motivation

The use of High Performance Computing (HPC) facilities allows for the implementation of computationally intensive statistical data assimilation techniques. We want to employ some of these techniques to assimilate temporal/spatial observations into a realistic one-dimensional ocean model. The model is based on partial differential equations (PDEs) and simulates physical and biogeochemical ocean properties.

Our goals:

- creating a biogeochemical model for the Bermuda region in the North Atlantic Ocean embedded within a ocean turbulence model,
- using ensemble-based statistical data assimilation techniques to incorporate physical, biological and chemical data into the model,
- making use of HPC techniques that allow an ensemble of model simulations and a data assimilation procedure to be run in parallel on an HPC cluster.

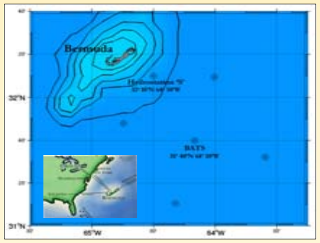
Observations

The Bermuda Atlantic Time-series Study (BATS) is a long-term monitoring program at a site in the western gyre of the Sargasso Sea, 50 nautical miles southeast of Bermuda. The observational data may be retrieved from the world-wide-web at http://bats.bbsr.edu/. For overview of the BATS study, see Steinberg et al. (2001).

Currently we examine the 183 available core cruises. These encompass the period from October 1988 to December 2003 with approximately one cruise per month. We consider measurements at depths ranging from 0 to 4300 m. Our variables of interest are physical and optical measurements from sensors like CTD, and biological and chemical variables measured on water samples that have been collected in Niskin bottles.

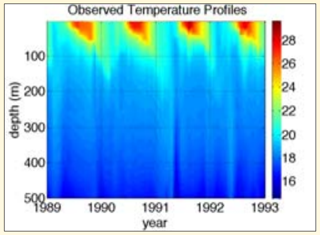
CTD data: 785,397 samples

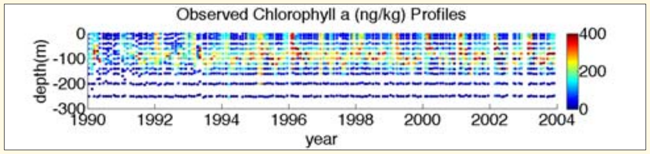
Variable	Units	Number of Values
Temperature	°C	785,397
Salinity	PSU	778,375
Oxygen	µmol/kg	732,636
Beam Attenuation Coefficient	1/m	170,904
Fluorescence	rel. fl. unit	520,359



Bottle data: 16,670 samples

Variable	Units	Number of Values
Temperature	°C	16,611
Salinity	PSU	6,265
Nitrate+Nitrite	µmol/kg	6,501
Particulate Organic Nitrogen	µg/kg	3,577
Chlorophyll a	ng/kg	2,015
Density Anomaly	kg/m³	16,083





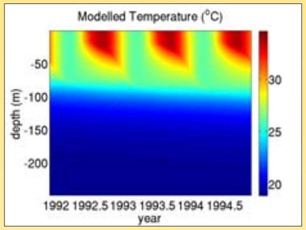
Physical Model

In order to simulate fluid motion and the effects of turbulent mixing on the distribution of the biological state variables we use the General Ocean Turbulence Model (GOTM, http://www.gotm.net). GOTM is written in Fortran 90/95. It includes different turbulence parameterizations as options for dynamic, spatially one-dimensional simulations of the upper ocean.

The core set of PDEs in GOTM describes transport of momentum, temperature and salinity as follows:

$$(1)\quad \frac{\partial u}{\partial t}-\frac{\partial}{\partial z}\left(K_t\frac{\partial u}{\partial z}\right)-fv=F_u \qquad (3)\quad \frac{\partial T}{\partial t}-\frac{\partial}{\partial z}\left(K_t'\frac{\partial T}{\partial z}\right)=F_T$$

$$(2)\quad \frac{\partial v}{\partial t}-\frac{\partial}{\partial z}\left(K_t\frac{\partial v}{\partial z}\right)+fu=F_v \qquad (4)\quad \frac{\partial S}{\partial t}-\frac{\partial}{\partial z}\left(K_t'\frac{\partial S}{\partial z}\right)=F_S$$



Here *u* and *v* are the current velocities in east-west and north-south directions, respectively. *T* denotes temperature and *S* salinity. The terms *fu* and *fv* denote the influence of the Coriolis force on current velocities. F_u and F_v represent pressure gradients and surface wind forcing. F_T and F_S are the sources and sinks for *T* and *S* due to heating and cooling and precipitation and evaporation at the sea surface. K_t is the turbulent viscosity for *u* and *v*, and K_t' is the turbulent diffusivity for *T* and *S*. GOTM includes a variety of turbulence model options to compute these diffusivities.

Biological state variables that are integrated into GOTM become subject to turbulent mixing as follows:

$$(5)\quad \frac{\partial X}{\partial t}-\frac{\partial}{\partial z}\left(K_t'\frac{\partial X}{\partial z}\right)=SMS(X)$$

Where $X \in \{Phy, Det, DIN, Chl\}$ denotes one of the biological state variables. *SMS(X)* models the sources and sinks of *X* and has the same functional form as the right hand side of the biological equations (6)-(9).

Biological Model

Our biological model is a simplified version of the model of Fennel et al. (2006) and includes 4 state variables in a system of PDEs: phytoplankton biomass (*Phy*), phytoplankton chlorophyll (*Chl*), detritus (*Det*) and dissolved inorganic nitrogen (*DIN*). The biological equations (6) – (9) describe the sources and sinks of our bio-chemical variables due to bio-chemical transformations. They are coupled with our physical model through equation (5) above.

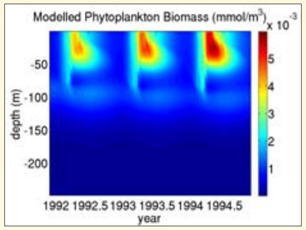
$$(6)\quad \frac{\partial Phy}{\partial t}=\mu Phy-m_P Phy-w_P\frac{\partial Phy}{\partial z} \qquad (8)\quad \frac{\partial DIN}{\partial t}=-\mu\, Phy+r_D Det$$

$$(7)\quad \frac{\partial Det}{\partial t}=m_P Phy-r_D Det-w_D\frac{\partial Det}{\partial z} \qquad (9)\quad \frac{\partial Chl}{\partial t}=\rho_{Chl}\mu Chl-m_p Chl-w_p\frac{\partial Chl}{\partial z}$$

Variable	Description	Units
Phy	Phytoplankton	mmol N m^{-3}
Det	Detritus	mmol N m^{-3}
DIN	Dissolved Inorganic Nitrogen	mmol N m^{-3}
Chl	Chlorophyll	mg m^{-3}
r_D	Remineralization rate	d^{-1}
w_D, w_P	Sinking Velocity	m/d

Here μ is a parameterization for photosynthetic growth of phytoplankton that depends in the availability of light and nutrients.

ρChl is a parameterization that accounts for photoacclimation, i.e. the synthesis of chlorophyll in order to optimize light harvesting for photosynthesis.



Data Assimilation

Idea: combine measurements and dynamical models to estimate model state and model parameters

Examples:
Optimize:
- biological parameters like phytoplankton growth and mortality
- dynamical terms like the vertical flux of organic matter
- initial or boundary conditions

Estimate and Forecast:
- the time evolving physical and biogeochemical state.

Framework: nonlinear and nonGaussian state space models

Approach: sampling-based solutions using sequential Monte Carlo methods, e.g. ensemble Kalman filter, particle filters (see Dowd 2007).

Computational Issues

Ensemble-based methods require many numerical model integrations (>10^2-10^3), hence, they are computationally expensive for realistic applications.

BUT: Ensemble-based methods should scale extremely well in massive parallel computing environments, especially when compared to iterative variational optimization methods (the most commonly used approach for assimilation in biological and physical ocean modelling at present).

Possible computational model: each of the 10^2 – 10^3 simulations are assigned to one CPU core.

Communication between cores is only necessary during analysis steps, when observations are ingested and the state of ensemble members is updated (i.e. assimilation steps).

References and Acknowledgments

Dowd, M. "Bayesian Statistical Data Assimilation for Ecosystem Models using Markov Chain Monte Carlo" *Journal of Marine Systems* (in press)

Fennel, K. and 5 others "Nitrogen cycling in the Mid Atlantic Bight and implications for the North Atlantic nitrogen budget: Results from a three-dimensional model" *Global Biogeochemical Cycles* 20 GB3007, doi:10.1029/2005GB002456 (2006)

Steinberg, D.K. and 5 others "Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): a decade-scale look at ocean biology and biogeochemistry" *Deep-Sea Research II* 48, 1405-1447 (2001)

PM and KF gratefully acknowledge funding from the ONR MURI program. KF is also supported through the CRC and NSERC Discovery programs. MD and JF are grateful for the support of NPCDS and NSERC.