Parameter optimisation of ocean biogeochemical models

Chloe Y. Huang | Supervisor: John R. Taylor | DAMPT (Ocean Dynamics Group)

Background

OceanBioME.jl is a fast and flexible ocean biogeochemical modelling environment written in Julia. It was designed specifically for ocean carbon dioxide removal applications, with flow physics, ecosystem models, and carbonate chemistry all seamlessly integrated within the same model. In order to model ocean processes realistically, the biogeochemical (BGC) models in OceanBioME.jl require several parameters - whose values change across different areas of the ocean - to be configured.

Due to the complex and often unmeasurable nature of these parameters and the large computational requirements of more complex BGC models, it may be difficult to estimate some of these parameters directly. Thus, it is often useful to perform parameter optimisation to better estimate these parameters and gain insight on the significance of a parameter on the model outcome. In the case of the BGC models, the output of the model comes in the form of several timeseries, which can then be further processed to produce statistics (such as mean, rms, and extreme values) to inform the features of the timeseries.

In short, for θ, y, where y is the observation and θ is the vector of parameters, we seek to solve problems of the form

$$y = \mathcal{G}(\theta) + \eta$$

where $\mathcal{G}$ is the forward map (the model output) and η is the (ideally Gaussian) noise.

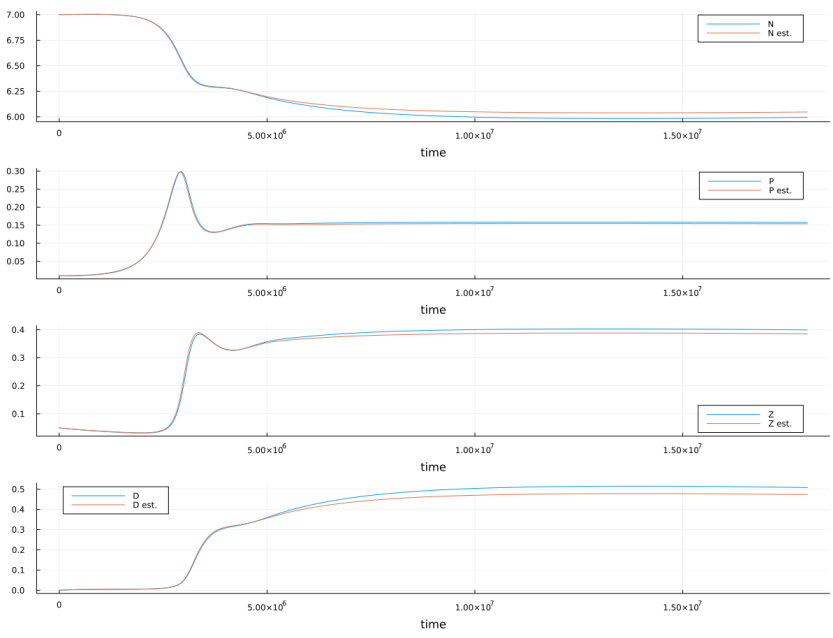
Process

Due to the nonlinear and non-Gaussian nature of the data being considered, Ensemble Kalman Filter (EnKF) approaches, often used in climate modelling, have been used. EnKFs are a class of derivative-free Bayesian optimisation techniques that use Monte Carlo sampling to estimate means and error empirically and are implemented in Julia via the EnsembleKalmanProcesses.jl package. Using this package, we produced a tool that allowed OceanBioME.jl models to be optimised given 'truth' data and prior guesses for the parameters and distribution of error in the observations.

The optimisation process was then tested over both artificial data, generated by a 'truth run' of the model with Gaussian noise added in post, and real GLODAP v2 data, which had error estimates for the measurements but otherwise no other information about the distribution of error. The artificial data was used to optimise BGC models, for which the target of optimisation were parameters like zooplankton mortality rate and phytoplankton chlorophyll ratio; the GLODAP data was used to optimise the CarbonChemistry model, for which the target of optimisation were the coefficients of the equations in the model itself.

Results

In general, optimisation was fairly successful for cases where the model output was not excessively sensitive to small changes in the parameters and the difference between prior guesses and true parameters were within 50% of the true values. However, optimisation was significantly less effective when the model was very sensitive or unstable due to the stochastic nature of EnKF processes.



Comparison between estimated and true timeseries of the NPZD model.

Hence, when trying to optimise the BGC models, optimisation was very successful, as seen in the figure on the left. The difference between the optimised and true parameters were almost always within 20% of the true parameters. However, in the case of the CarbonChemistry model, which is not only very sensitive to small changes but also optimised over relatively noisy data, EnKF processes were not as easily implemented.

Under current methodology, the instability of the model has overcome the potential success of ensemble based approaches like EnKFs, resulting in only minor improvements from the original model. However, attempts at optimising with the current methods have helped inform correlations between some parameters and model error. More statistical analysis and potentially new approaches are required to achieve better results.