

1 Context

Objectives: I want to predict the nutrient flux from the deep sea to the surface layer in the northern Red Sea. This is interesting, because it could give us an idea about the new production. With existing methods, this amount of new production is difficult to estimate without in-situ data. Therefore if there is a method that allows us to do that by just using satellite data, it could be of interest for marine biologist.

Data: I clustered the red sea into 8 provinces bases on monthly chl data and a Gaussian mixture model clustering algorithm. I am interested in the northernmost cluster during winter 1999-2000 because there is an important bloom in this region. I averaged daily SeaWifs data from December 1999 to May 2000 in the given cluster, giving me the following a time series data.

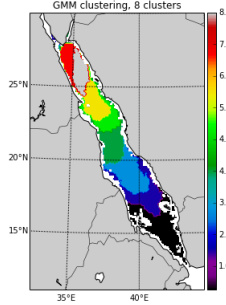


Figure 1: Result of clustering. We are interested in the red region.

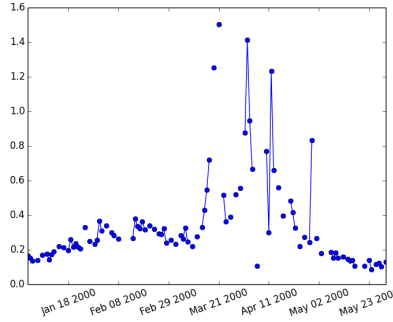


Figure 2: Averaged of CHL concentration in studied cluster: The data to be used in the inversion.

Box model: We assume that a box model or 0d model represents the mixed layer in the cluster region we are interested in, and furthermore that there is no exchange with the neighboring regions and that the input of nutrient just come from the deep sea. We only model the phytoplankton, the nutrient concentration (we assume that the nitrate is the limiting nutrient), and the grazing pressure through a zooplankton concentration. A certain part of the zooplankton and phytoplankton is assumed to be lost by sinking, whereas a stochastic term models a flux of nutrient that diffuse to the mixed layer from the deep water. This is that term that we are interested in estimating.

NPZ model: The NPZ model is the simplest meaningful model for the phytoplankton growth. It is a ODE model that links phytoplankton concentrations (P) to preying zooplankton (Z) and available nutrient (N) concentrations. In our case we augment the NPZ model by including a nutrient flux term (Φ) that models an input of nutrient from the deep sea. The equations of the NPZ model are shown below:

$$\frac{dP}{dt} = \mu \frac{N}{k + N} P - gPZ - \varepsilon_P P - \varepsilon_P^* P^2, \quad (1)$$

$$\frac{dZ}{dt} = \gamma gPZ - \varepsilon_Z Z - \varepsilon_Z^* Z^2, \quad (2)$$

$$\frac{dN}{dt} = -\mu \frac{N}{k + N} P + (1 - \gamma)gPZ + \varepsilon_P P + \varepsilon_Z Z + \Phi_N(t). \quad (3)$$

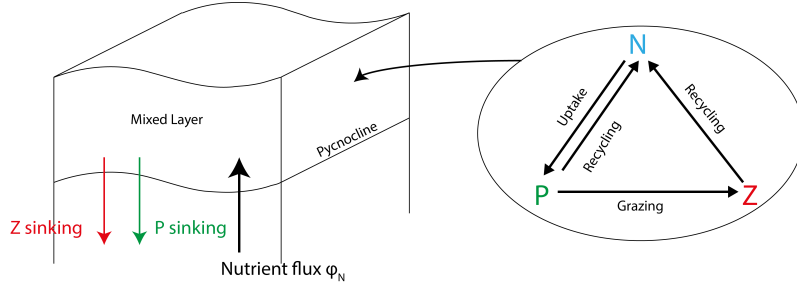


Figure 3: Box model

2 State-Space model

Denote Y_1, \dots, Y_T the observations of chlorophyll concentration, $\mathbf{X}_1, \dots, \mathbf{X}_T$ the state variables, with $\mathbf{X}_t = [N_t, P_t, Z_t, \Phi_t]^T$, the concentration of each of the ODE state variables at time t . For now on, we actually consider the log of these concentrations.

State Equation: The state variables are assumed Markovian:

$$X_t = \begin{bmatrix} N_t \\ P_t \\ Z_t \\ \Phi_t \end{bmatrix} = \begin{bmatrix} f(X_{t-1}; \theta) \\ \alpha \Phi_{t-1} \end{bmatrix} + \mathbf{w}_t, \mathbf{w}_t \sim \mathcal{N}(0, \text{diag}(\Gamma)) \quad (4)$$

f is the result of the integration of the ODE system between times $t-1$ and t with initial conditions $[N_{t-1}, P_{t-1}, Z_{t-1}]$ and a fixed nutrient flux Φ_{t-1} . Γ is the noise model covariance. θ are the NPZ model parameters. Φ_t is an autoregressive process with coefficient $\alpha = 0.7$.

Observation equation: The observations are independent conditionally on the state of the system.

$$Y_t = c + P_t + w_t^o, \quad (5)$$

with $w_t^o \sim \mathcal{N}(0, (\sigma^o)^2)$ iid and c in the log-conversion rate between the chlorophyll and the phytoplankton concentrations.

3 EM methodology

$\Theta = \{\Gamma, \sigma_o^2, x_0^b, B\}$ are unknown and estimated by with the iterative Expectation-Maximization algorithm, with latent variable the states $X_{0:T}$. In the E-step, the distribution $p(X_{0:T}|\Theta, Y_{0:T})$ is estimated given the current estimate of the parameters Θ . Since it is a smoothing problem with a non linear model, I use the Ensemble Kalman Smoother. In the M-step, Θ is updated by maximizing $\log p(\Theta|X_{0:T}, Y_{0:T})$

4 Problems

- The model needs to be validated by someone who knows about ecological modelling and the biology of the Red Sea. Currently I don't know:
 - If a 0d model is appropriate
 - If the NPZ captures enough of the dynamics
 - What should be the transfer function and the parameters
 - If the stochastic nutrient flux assumption is reasonable
- With the current model I run frequently into numerical instability issues
- The parameter values are unknown, we tried a state-parameter estimation, but it doesn't work except is we start around the true value and put a ridiculous amount of noise.