

1 Context

Objectives: Here I want to predict the nutrient flux because it can give us an idea about new production. This amount of new production is difficult to estimate without data at high temporal resolution with existing methods [aqaba paper]. Therefore if there is a method that allows us to do that just using satellite data that would be great.

Data: I clustered the red sea into 8 provinces bases on monthly chl patterns. [cluster figure] I want interested in the cluster nb ? during winter 1999-2000 because there is an important bloom in this region. I averaged daily SeaWifs data from [] to [] in the given cluster, giving me the following time series [figure], that is to say [] nb of measurements.

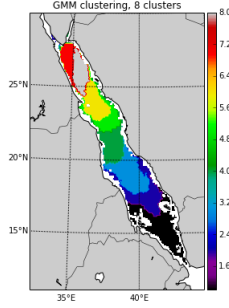


Figure 1: Result of clustering. We are interested in the red region.

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:

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

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

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

2 Statistical modeling

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 considere the log of these concentrations.

Data model: The observations are independent conditionally on the state of the system.

$$Y_t = c_t + P_t + w_t^o, \quad (4)$$

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

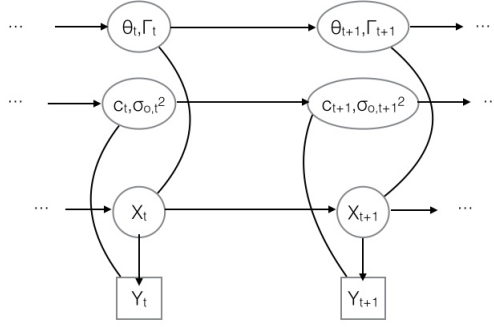


Figure 2: BHM second attempt

Process model: The state variables are assumed Markovian:

$$X_t = \begin{bmatrix} N_t \\ P_t \\ Z_t \\ \Phi_t \end{bmatrix} = \begin{bmatrix} f(X_{t-1}; \theta_t) \\ \Phi_{t-1} \end{bmatrix} + \mathbf{w}_t, \mathbf{w}_t \sim \mathcal{N}(0, \text{diag}(\Gamma_t)) \quad (5)$$

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} . $\Gamma_t = [(\sigma_t^N)^2, (\sigma_t^P)^2, (\sigma_t^Z)^2, (\sigma_t^\Phi)^2]$. Φ_t is a random walk process.

Transition model: We denote: $\lambda_t = [\theta_t^T, \Gamma_t^T, c_t, \sigma_t^o]^T$

We have to specify a transition model for the time-varying parameters. We choose a random walk process of the form:

$$\text{logit}(\lambda_{t+1}^{(i)}) = \text{logit}(\lambda_t^{(i)}) + v_t^{(i)}, \text{ if } \lambda_t^{(i)} = \gamma_t, \quad (6)$$

$$\log(\lambda_{t+1}^{(i)}) = \log(\lambda_t^{(i)}) + v_t^{(i)}, \text{ otherwise,} \quad (7)$$

where v_t^λ is a iid Gaussian random error of fix variance.

3 EM methodology

4 Case where θ is known

5 Case where θ is unknown